GRAMMATICAL CONSTRAINTS IN SECOND LANGUAGE SENTENCE PROCESSING

BY

EUNAH KIM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
with a concentration in Second Language Acquisition and Teacher Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Silvina Montrul, Chair
Professor James Hye-Suk Yoon
Associate Professor Tania Ionin
Associate Professor Kiel Christianson

Abstract

A central issue in L2 research concerns the nature of grammatical representations that late L2 learners come to develop in the L2. Previous work suggests that L2 learners sometimes underuse morpho-syntactic information during online processing of L2 sentences, leading to an ongoing debate about how they represent and process structural information in sentence processing. Some researchers propose that L2 sentence processing is qualitatively different from L1 sentence processing in that the former characteristically involves ‘shallow’ structural analysis (Clahsen & Felser, 2006a), whereas other researchers suggest that the differences between L1 and L2 processing are attributable to quantitative factors such as the amount of language experience, the proficiency level in the target language or the availability of processing resources (Frenck-Mestre, 2002; Hopp, 2006; McDonald, 2006). The present dissertation seeks to further our understanding of adult L2 syntactic processing by examining L2 learners’ sensitivity to ‘island constraints’ (Ross, 1976) in the course of online processing of long-distance wh-dependencies, using plausibility judgments and eye-movement monitoring techniques.

In Experiment 1, a stop-making-sense task was conducted to investigate L2 learners’ sensitivity to the subject/relative clause island constraint in online plausibility judgments. The native speakers showed immediate sensitivity to island constraints, as evidenced by the fact that although they interpreted a wh-dependency at the earliest possible gap site when it is grammatically licit, they suspended the immediate gap postulation within a syntactic island. The L2 learners were not as efficient as native speakers in suppressing active gap search, but they ultimately ruled out an illegal dependency, in accordance with the island constraint.

Experiment 2 employed eye-movement monitoring techniques to examine the way native speakers and L2 learners apply the subject/relative clause island constraint when processing

filler-gap dependencies under a more natural reading situation. Working memory capacity of the participants was also measured in an attempt to capture potential individual differences in filler-gap processing and grammar application. The results indicate that even native speakers failed to immediately suppress automatic active gap creation inside an island in this natural reading situation, and both groups applied island constraints at a later stage. There was also suggestive evidence that readers with larger working memory capacity applied island constraints earlier than those with smaller working memory capacity, in both native speakers and L2 learners, suggesting that more processing resources may allow a more rapid and efficient application of grammatical constraints.

Experiment 3 investigated whether L2 learners are sensitive to a more subtle grammatical distinction—differential distribution of parasitic gaps within two kinds of extraction islands (i.e., subjects with an infinitival complement vs. subjects with a finite relative clause modifier). The native speakers' eye movement patterns showed evidence for a rapid distinction between the two types of islands, indicating active gap search only in the island that allows a parasitic gap. Some of the L2 learners showed a similar pattern of sensitivity to this subtle grammatical constraint, though the effect appeared in a later region as compared to the native speakers.

The three experiments showed additional quantitative differences between the native speakers and L2 learners, including differences in reanalyzing the incorrect initial wh-dependency and in speed of constructing a complex phrase.

Taken together, these findings suggest that L2 learners are sensitive to an array of important linguistic and cognitive factors in ways that are *qualitatively* comparable to native speakers in online sentence processing, and their comprehension is based on detailed, grammatically-correct syntactic parse of L2 sentences. The L1-L2 differences in sentence

processing can be largely attributed to *quantitative* factors such as processing speech/efficiency or working memory capacity differences. The findings are discussed in terms of the theoretical debate on the nature of the L2 sentence processing mechanisms.

to Bang Yeol Kim and Yeon Im Cheong, my parents

Acknowledgements

I would like to express my deepest thanks to my advisor Silvina Montrul, and the members of my dissertation committee, James Hye-Suk Yoon, Tania Ionin and Kiel Christianson for their valuable comments on this dissertation and for their warm and constant support throughout my doctoral study. My heartfelt thanks go to my husband and my best friend, Soondo Baek, who has always been with me in good and bad days. Finally, I thank National Science Foundation for Doctoral Dissertation Research Improvement Grants (Award Number: 1022608, PI: Silvina Montrul, co-PIs: Eunah Kim, James Hye-Suk Yoon, Kiel Christianson), which allowed me to run Experiments 2 and 3 of this dissertation and to present the results in conferences. This dissertation was able to be completed with all these helps.

Table of Contents

Chapter 1

Introduction

Comprehending a sentence as it is coming into the listener or reader involves complex processes responsible for the incorporation of various kinds of information provided by the sentence. Information at multiple levels of linguistic description, including phonological, lexical, syntactic, semantic, pragmatic and discourse information, needs to be accessed rapidly and incorporated into the developing sentence representation in order to arrive at the correct understanding of the incoming sentence. For second language (L2) learners of a language, rapid access to and coordination of manifold language cues should be much more demanding than for native speakers of that language. Consequently, the products of L2 sentence processing may well be different in many respects from those of native language sentence processing.

Research on L2 phenomena has recently paid growing attention to the question of how L2 learners comprehend sentences in real-time. Psycholinguistic research techniques that provide time-sensitive information have been adopted fruitfully to investigate the characteristics of the representations and processes that L2 learners employ to comprehend incoming input on a moment-by-moment basis, and what kinds of information from the input they are sensitive or insensitive to during online sentence interpretation (e.g., Frenck-Mestre, 2005; Juffs, 2001; Marinis, 2003; Mueller, 2005; Papadopoulou, 2005; Roberts, 2012). This line of research contributes to enhancing our understanding of the L2 processing mechanisms and provides valuable insights into how L2 learners develop their knowledge in the L2 at various levels of language (Carroll, 1999)

As research on L2 acquisition has resorted to the native speakers' linguistic competence to understand the characteristics of the L2 knowledge, studies on L2 sentence processing have characterized L2 learners' performance in online L2 processing by comparing it with that of native speakers. This approach has been motivated by the question of whether the L2 processing mechanisms are of the same kind as the L1 processing mechanisms, in other words whether L2 learners can develop native-like processing routines. Studies conducted so far have begun to pinpoint linguistic levels in which native speakers and L2 learners converge or diverge in terms of their sentence processing behaviors. The overall results suggest that L2 learners can make use of lexical, semantic, pragmatic, and discourse-level information in ways that are very similar to how native speakers process those kinds of information (e.g., Dussias & Pinar, 2010; Felser, Roberts, Marinis & Gross, 2003; Frenck-Mestre & Pynte, 1997; Pan & Felser, 2011; Robert & Felser, 2011; Williams et al., 2001). By contrast, studies on L2 learners' ability to use morpho-syntactic information have produced inconsistent results—some studies have found that L2 learners are able to conduct structural analysis of input sentences as efficiently and accurately as native speakers, whereas other studies have found that syntactic representations constructed by L2 learners are not as complete as what native speakers build during online sentence processing (e.g., Bowden, Steinhauer, Sanz & Ullman, 2013; Felser & Roberts, 2007; Hopp, 2006, 2010; Marinis, Roberts, Felser & Clahsen, 2005).

These conflicting findings drive the debate on the nature of L2 syntactic processing, which has been one of the central issues in L2 sentence processing research. On one end of the different views on that matter, Clahsen and Felser (2006a) propose that L2 learners' processing of syntactic information is deficient as compared to native speakers of the target language. According to their hypothesis (the Shallow Structure Hypothesis), while sentence processing by native speakers is typically based on fully specified syntactic representations (cf, see Christianson, Hollingworth, Halliwell & Ferreira, 2001; Ferreira & Patson, 2007; Sanford & Sturt, 2002 for evidence otherwise), adult L2 learners are unable to compute as detailed syntactic representations out of the incoming input. L2 sentence processing is thus characterized as limited to 'shallow' structural analysis of a given sentence, consequently being led to heavy reliance on extra-syntactic information such as lexical, semantic and pragmatic information. On the other end, there are proposals that can be broadly subsumed as the 'quantitative' view of the differences between L1 and L2 sentence processing. According to this view, although adult L2 learners of advanced proficiency in the L2 sometimes appear to be unable to construct fully specified syntactic representations of L2 sentences, it does not necessarily imply that there is a categorical boundary that qualitatively distinguishes between syntactically complete L1 parses and 'shallow' L2 parses. The differences in the degree of the completeness with which native speakers and L2 learners analyze the syntactic structure of a sentence are instead seen to be due to some quantitative factors such as the amount of language experience, the availability of processing resources, and the proficiency level in the target language (e.g., Frenck-Mestre, 2002; Hopp, 2006; McDonald, 2006). Frenck-Mestre (2002), for example, suggests that L2 learners can incorporate syntactic-structural information in native-like ways, provided that they have sufficient amount of exposure to and/or high enough proficiency level in the L2. McDonald (2006) has sought explanations for poor grammatical processing by L2 learners from their less efficient processing of low-level linguistic information such as lexical perception and word retrieval, rather than from deficits in grammatical representations. Under this view, limitations on processing resources on the part of L2 learners, which may well be aggravated by the need to communicate in a non-native language, could make it difficult for L2 learners to access and use grammatical information even when they have acquired the relevant linguistic knowledge.

In line of this debate on the nature of L2 sentence processing mechanisms and underlying grammatical representations, the present dissertation seeks to further our understanding of adult L2 syntactic processing by examining the ways L2 learners do or do not use ‘island constraints’ in the course of online processing of wh-questions. Island constraints refer to restrictions on movement out of certain syntactic domains, collectively called ‘islands’. For example, extraction of a wh-phrase out of a relative clause is deemed ungrammatical in English as shown in example (1).

(1) *How many cities does John have brothers [who live in ___]? (relative clause island)

Since the seminal work of Ross (1967), various kinds of constructions that block syntactic extraction have been identified, and much research has been conducted to uncover rules that govern the island phenomenon. This work suggests that there is a set of abstract syntactic constraints underlying the island phenomenon, which operate over hierarchically organized syntactic structures (e.g., Chomsky, 1986; Rizzi, 1990).

Previous studies on sentence processing by monolingual native speakers revealed that island constraints operate immediately in real time sentence processing to prevent the formation of illicit wh-dependencies (Bourdages, 1992; Kluender & Kutas, 1993; McElree & Griffith, 1998; McKinnon & Osterhout, 1996; Phillips, 2006; Picking, Barton, & Shillcock, 1994; Stowe, 1986; Traxler & Pickering, 1996). Assuming that the island phenomenon is best explained by grammatical constraints that operate on hierarchical syntactic configurations, the previous finding of the online sensitivity to the island constraints suggests that the native language sentence processing architecture is designed in such a way that fairly detailed syntactic relations among the words in a sentence is incrementally constructed and evaluated. This observation motivated the present dissertation, which aims to test English L2 learners' sensitivity to English island constraints during the course of online wh-dependency formation.

Investigation of adult L2 learners' sensitivity to island constraints during online processing of wh-dependencies can offer valuable implications for understanding the characteristics of L2 sentence processing and evaluating the relevant previously proposed hypotheses. As mentioned above, the operation of island constraints presupposes the construction of complex syntactic configurations that constitute island domains. Therefore, if L2 learners show immediate sensitivity to island constraints during online L2 processing, we could infer that they were successful in computing as detailed syntactic representations as required by the constraints to operate. This bears on the debate over the specificity of the syntactic representations constructed by L2 learners.

Moreover, the question of whether L2 learners can apply complex grammatical constraints during online sentence processing is in itself an interesting question in L2 processing research, because it provides an important criterion by which L1 and L2 sentence processing mechanisms can be compared. It also bears on the question of how the L2 grammatical constraints are represented by adult L2 learners. Are they represented as some kind of explicit/declarative knowledge that does not allow rapid access during online processing? Or can they be turned into implicit/procedural knowledge that can be readily accessed in sentence comprehension? The present dissertation experiments were designed to shed light on these questions.

This dissertation can also offer insights regarding the issue of accessibility to Universal Grammar (UG) in adult L2 acquisition, which has been a major issue in the traditional L2 acquisition research (e.g., Bley-Vroman, 1990, 2009, Schwartz & Sprouse, 1996). Under the assumption that island constraints are part of UG, the learnability of the constraints has received a significant amount of attention in the context of the UG-accessibility debate (e.g., Bley-Vroman, Felix, & loup, 1988; Hawkins & Chan, 1997; Johnson & Newport, 1991; White & Juffs, 1998). Previous research, however, mostly relied on offline grammaticality judgment tasks for data, which provide information primarily on how island constraints influence participants' final judgments. As this task has limited use in revealing the processes through which the participants reach final judgments, research on the online application of island constraints would be able to provide us the missing part of information on the state of knowledge of island constraints in L2 learners, regarding which the present study can contribute to the literature of L2 acquisition.

This dissertation reports three experiments that aimed to investigate the ways adult L2 learners process wh-dependencies and the ways they apply island constraints in real time comprehension. Experiment 1 employed an online plausibility judgment task to probe into L2 learners' potential to apply the island constraints. Experiments 2 and 3 used eye-movement monitoring to investigate L2 learners' processing of wh-dependencies and sensitivity to the island constraints under natural reading situations. Considering the complexity of island constraints, the dissertation targeted learners with a fairly high L2 proficiency. The learner participants were Korean-speaking learners of English. Korean is a wh-in-situ language, for which previous L2 research on acquisition of island constraints assumed that the island constraint does not operate (e.g., Schachter, 1990, but see Chapter 3 of this dissertation for more discussion). The dissertation tested learners who were immersed in the L2-speaking environment after puberty, because theoretically important issues in the L2 processing of complex syntactic structures (e.g., Shallow Structure Hypothesis) and L2 acquisition of island constraints (e.g., UG accessibility debate) mainly concern adult L2 learners who have begun to learn the L2 after the offset of the Critical Period (Lenneberg, 1967) in language acquisition (e.g., Bley-Vroman, 1990; Clahsen & Felser, 2006a; Johnson & Newport, 1989).

The dissertation is organized as follows. Chapter 2 presents a review of the debate around the nature of L2 syntactic processing. Chapter 3 discusses how the island phenomenon has been dealt with in the syntax and processing literature and summarizes the findings of relevant previous L2 studies. Chapter 4, 5 and 6 report the results of three experiments conducted to examine the role of island constraints in L2 processing. The findings from these experiments are discussed in Chapter 7.

Chapter 2

Sentence processing in a second language

2.1. The Shallow Structure Hypothesis

One of the main concerns in L2 research has been to characterize the grammatical knowledge of L2 learners. Numerous studies have investigated the knowledge that L2 learners have about various grammatical phenomena (See Hawkins, 2001; White, 2003 for reviews). Compared to the history of the research on L2 grammatical knowledge, the history of research on the use of this knowledge in real-time sentence comprehension is relatively short. There is a growing interest in the real-time processes in which L2 learners comprehend sentences. How L2 learners construct a mental representation of the sequentially incoming language input and what types of information the learners use in the process is not only an interesting topic in itself, but also an important topic in that it provides us a more complete picture of the language of the learners.

The leading issue of L2 sentence processing research has been whether L2 processing is qualitatively different from L1 processing, or whether L1 and L2 processing are fundamentally similar (e.g., Clahsen & Felser, 2006a, b,c; Dekydtspotter, Schwartz & Sprouse, 2006; Frenck-Mestre, 2002). This is the processing version of the long-standing debate in L2 research over the cause of the persistent problems with morpho-syntactic representation experienced by post-puberty learners. The difficulty has been viewed as evidence for a permanent loss of ability to acquire a native-like grammatical representation (e.g., Bley-Vroman, 1990; Hawkins & Hattori, 2006), or alternatively, as a problem with retrieving relevant grammatical knowledge in real time (e.g., Goad & White, 2006; Prévost & White, 2000).

An influential view on the nature of second language processing is embodied in the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, b). Based on empirical evidence suggesting that adult L2 processing does not make use of full range of syntactic information (e.g., Felser, Roberts, Gross & Marinis, 2003; Marinis, Roberts, Felser & Clahsen, 2005), it was proposed that L2 learners resort to 'shallow' syntactic processing. In the original Shallow Structure Hypothesis (Clahsen & Felser, 2006a), shallow syntactic processing was defined as failing to compute hierarchical relations between words and failing to build representations of certain syntactic elements such as empty categories. Clahsen and Felser (2006b) and Felser and Roberts (2007) introduced Hammerton, Osborne, Armstrong and Daelemans's (2002) definition of shallow parsing and suggest that shallow parsing typically involves identifying parts of speech, segmenting the input string into meaningful chunks (i.e., phrasal or clausal units), and determining what relations (e.g., subject, object, etc.) these chunks bear to the main verb. Further elaborating on the concept of shallow processing, Clahsen and Felser (2006c) suggested that L2 syntactic processing can be fully conducted in a local domain, but what causes difficulty for even highly proficient L2 learners is to process non-local dependencies.

According to the hypothesis, the adult L2 learners' reduced ability to conduct full syntactic analysis is "due to the knowledge source that feeds the structural parser, the L2 grammar, being incomplete, divergent, or of a form that makes it unsuitable for parsing." (Clahsen & Felser, 2006b, p. 117). Adopting the view that adult L2 learners' interlanguage grammar is fundamentally different from native speakers' grammar (e.g., Bley-Vroman, 1990; Clahsen & Muysken, 1986, 1989), the hypothesis suggests that sufficiently detailed syntactic parsing is limited in L2 processing. L2 processing consequently relies heavily on the information that it can access (e.g., lexical, semantic, pragmatic or discourse information), as well as rough and ready, semantically-based comprehension heuristics (Townsend & Bever, 2001). L2 processing is thus sharply contrasted with L1 processing, which is assumed in this account to incorporate full syntactic information quickly and accurately.

An important claim of this hypothesis is that differences between L1 and L2 processing cannot be explained in terms of quantitative variables. Differences between native speakers and L2 learners, and among learners, in their online comprehension processes are not attributed to differences in proficiency in the target language, amount of exposure to the target language, available processing resources or processing speed. Moreover, properties of learners' native language do not influence learners' processing patterns. In other words, shallow processing is not restricted to a certain class of learners (e.g., those who have a low proficiency/little exposure/limited processing resources/an L1 that is different from the target language at a certain grammatical level), but is a characteristic property of all adult L2 learners. Differences between L1 and L2 processing are therefore viewed as qualitative, not quantitative, in nature.

The idea of the Shallow Structure Hypothesis that adult L2 processing relies on shallow syntactic parse and is qualitatively different from L1 processing has attracted considerable attention and has been under debate in L2 research. As reviewed in the following section, existing studies have produced evidence both supporting and refuting the view that L2 processing underuses morpho-syntactic information for an architectural reason. Before discussing the studies, it would be worth noting the growing body of literature suggesting that even processing in one's native language is not always accurate and is sometimes quite shallow (e.g., Christianson et al., 2001; Christianson, Luke & Ferreira, 2010; Ferreira, 2003; Ferreira & Patson, 2007; Sanford & Sturt, 2002). For example, in Christianson et al. (2001), native English speakers read garden-path sentences such as *While Anna dressed the baby played in the crib*. and were given questions such as *Did Anna dress the baby*?. Although the correct answer to the question is ‘no’, participants answered ‘yes’ quite often and confidently, suggesting that they did not fully recover from the initial misanalysis. Ferreira (2003) showed that not only native speaker comprehenders misinterpret garden-path sentences, but they also sometimes misinterpret unambiguous sentences. Native English speakers misinterpreted implausible, non-canonical (i.e., passive) sentences such as *The cat was chased by the mouse*. more often than the plausible counterpart (e.g., *The mouse was chased by the cat*.), whereas they interpreted canonical (i.e., active) sentences accurately regardless of plausibility (e.g., *The mouse chased the cat*. vs. *The cat chased the mouse*.). This finding was interpreted as indicating that native English speakers employ heuristic processing mechanisms such as semantic plausibility and a Noun-Verb-Noun word order template that maps onto Agent-Verb-Patient interpretations (Townsend & Bever, 2001). The growing number of evidence that native speakers’ processing is shallow under certain circumstances raises the questions of whether drawing a categorical distinction between L1 and L2 processing in terms of ‘depth’ of syntactic processing is tenable, and if so, in what ways shallow processing in L1 and L2 are different—questions that need to be considered in the evaluation of the Shallow Structure Hypothesis.

The next section presents a review of major experimental studies that investigated L2 learners’ morpho-syntactic processing.

2.2. Use of morpho-syntactic information in L2 sentence processing

There is an increasing interest in what types of information L2 learners use in real time sentence comprehension and whether they use them in a native-like way. Although more work is needed to arrive at full understanding of the ways different types of information are used in L2 sentence processing, available studies generally suggest that proficient L2 learners are capable of incorporating various types of information in online processing. Learners have been shown to be sensitive to lexical-level information such as thematic roles (e.g., Felser et al., 2003; Papadopoulou & Clahsen, 2003) and verb-specific argument structures (Dussias & cramer Scaltz, 2008; Juffs, 1998, 2004; Frenck-Mestre & Pynte, 1997; Rah & Adone, 2010). They were also found to be sensitive to plausibility information (e.g., Felser & Roberts, 2011; Frenck-Mestre & Pynte, 1997; Roberts & Felser, 2011; Williams, Möbius, & Kim, 2001) and referential context (Pan & Felser, 2011). However, existing studies have produced mixed findings with regard to the use of information at the morpho-syntactic level, with some suggesting that even proficient L2 learners' sentence processing does not fully make use of morpho-syntactic information, whereas others show evidence of native-like use of the information. These studies are discussed in more detail below.

Behavioral studies

A relevant piece of findings come from studies investigating syntactic ambiguity resolution in L2 processing. Several studies examined the way L2 learners resolve the relative clause attachment ambiguity illustrated in (1).

(1) a. Someone shot the servant of the actress who was on the balcony.

b. Someone shot the servant with the actress who was on the balcony.

In (1a), with complex genitive NPs, the relative clause *who was on the balcony* can modify either *the servant* (NP1 attachment) or *the actress* (NP2 attachment). Previous L1 research has uncovered cross-linguistic differences in the attachment preference, with languages such as English, Brazilian Portuguese and Italian favoring NP2 attachment and languages such as Spanish, German, Dutch and Russian preferring NP1 attachment. In (1b), where a thematic preposition *with* introduces the second NP, a universal preference for NP2 attachment was found. Gibson, Pearlmutter, et al. (1996) accounts for the cross-linguistically different attachment preferences in sentences such as (1a) by assuming different relative strength of two phrase-structure-based locality principles. The principle of *recency*, according to which new incoming materials are attached to the most recently processed phrase, causes preference for NP2 attachment. Preference for NP1 attachment occurs because the principle of *predicate proximity*, according to which new materials should be attached as structurally close as possible to the head of a predicate phrase, outranks the principle of *recency* (cf, Frazier & Clifton, 1996; Mitchell & Cuetos, 1991 for different explanations for cross-linguistic differences in attachment preference). The universal NP2 attachment preference in sentences with a thematic preposition (1b) has been considered as suggesting a strong influence of thematic domain that overrides any phrase-structure-based locality principle that might otherwise favor NP1 attachment (Frazier & Clifton, 1996).

The relative clause attachment preferences in L1 processing were used to test whether L2 processing involves phrase-structure-based parsing principles and whether it is sensitive to thematic information. Moreover, since there are cross-linguistic differences in attachment preferences, the phenomenon was suitable to examine potential transfer from the L1 parsing strategy. Felser et al. (2003) and Papadopoulou and Clahsen (2003) examined what attachment preference advanced L2 learners have when processing sentences such as (1a-b) using self-paced reading. In Felser et al. (2003), the learners' L1s were German and Greek (languages with NP1 attachment preference) and the L2 was English (a language with NP2 attachment preference). In Papadopoulou and Clahsen (2003), both learners' L1s and the L2 were languages that prefer NP1 attachment (L1: Spanish, German, or Russian, L2: Greek). In both studies, native speakers exhibited the expected preference patterns when processing sentences with complex genitive NPs such as (1a) whereas L2 learners showed no preference for either attachment. When processing sentences with complex NPs linked by thematic prepositions such as (1b), both native speakers and L2 learners showed preference for NP2 attachment. The studies concluded that although L2 learners use lexical-semantic information to attach ambiguous relative clauses, the ambiguity resolution is not guided by phrase-structure information. They also concluded that the preference of a parsing principle in the learners' L1 does not affect L2 processing. These results were taken as prime evidence for the Shallow Structure Hypothesis and for (over-)reliance on extra-syntactic information in L2.

However, there are also studies showing that L2 learners do show relative clause attachment preferences when processing sentences with a complex genitive NP. Frenck-Mestre (1997), for example, tested English learners of French using eye-tracking. The learners were relatively non-proficient/had not had much exposure in French (a mean of three years of formal learning of French in a classroom setting outside of France, nine months of immersion in French in France, and roughly 5 on a 10 point scale of self-rated proficiency in French). Although the native French speakers showed an NP1 attachment preference in their first pass gaze durations, the learners showed an NP2 attachment preference in the same measure, which was the preference found in their native language. When the same task was given to a more proficient/more experienced English learners of French in Frenck-Mestre (2002) (roughly three years of formal learning of French outside of France, at least two years of study in a French university, a mean of five years of stay in France, and 7 or better on a ten point scale of self-rated French proficiency), the learners showed a native-like NP1 attachment preference. Together, the results suggest that relative clause attachment preference in L2 could be influenced by the preference in the learners' L1 at an early stage of learning, but the learners may come to use native-like attachment preference as they become more proficient/have more exposure in the L2. Crucially, the fact that both the lower- and the higher-level learners showed an attachment preference provides evidence against the view that the learners do not compute hierarchical relations in L2 processing. The results also suggest that proficiency/exposure can potentially influence L2 syntactic processing (see Dussias & Sagarra, 2007, for evidence that exposure in L2 influences relative clause attachment preference in bilinguals' L1 when L2 proficiency is controlled).

Another set of studies has investigated the depth of syntactic processing in L2 by examining whether the learners construct abstract syntactic elements such as gaps (traces) in filler-gap sentences. Using a cross-modal picture priming task, Felser and Roberts (2007) examined the way advanced Greek-speaking learners of English interpret filler-gap sentences, focusing on whether filler-gap processing is structurally mediated through a gap. The results were compared with those of Roberts, Marinis, Felser and Clahsen (2007), who tested native English speakers with the same task. Participants listened to sentences containing an indirect-object relative such as (2). At [2] (gap position) or at [1] (pre-gap position), which is an earlier control position, participants saw pictures visually presented on the computer screen and judged whether what the picture describes is alive or not. The visual targets were either identical to the indirect object noun (a squirrel for example (2)) or unrelated to it (toothbrush for example (2)). If participants construct a gap at [2] and if filler-gap processing mediated by the gap (i.e., the filler is reactivated at the gap position), then it was predicted that the response time of the aliveness judgment task would be shorter for the pictures identical to the filler than to the unrelated pictures at [2] but not at [1].

(2) Fred chased the squirrel to which the nice monkey explained the game's [1] difficult rules [2] in the class last Wednesday.

Adult native speakers with high working memory spans showed a speed advantage for the identical picture at the gap position but not at the pre-gap position, suggesting that high-span native speakers mentally reactivate the filler at the gap position. Native speakers with low working memory spans, however, did not show any facilitation for identical pictures at either of the two test positions. The reaction time pattern of Greek learners of English was different either from the high-span native speakers or from the low-span native speakers in that the learners' response time was faster for identical pictures than for unrelated pictures in both test positions. This position-non-specific facilitation effect in the L2 learner group was interpreted as evidence of maintained activation, not of reactivation at the gap site. Felser and Roberts concluded that L2 learners do not compute gaps when processing filler-gap sentences, but filler-gap sentences are comprehended in L2 by directly associating the filler and the subcategorizer.

Marinis et al.'s (2005) self-paced reading study investigated whether upper intermediate or advanced learners of English with diverse L1 background (Chinese, Japanese, German and Greek) make use of intermediate gaps when processing long-distance filler-gap sentences such as (3a), modeled on an L1 processing study by Gibson and Warren (2004).

(3) a. The nurse who the doctor argued $\_\_\__{IG}$ that the rude patient had angered ___ is refusing to work late. (+extraction, VP(+intermediate gap))

b. The nurse who the doctor's argument about the rude patient had angered ___ is refusing to work late. (+extraction, NP(-intermediate gap))

The filler-gap dependency spans across an intermediate gap site ($\_\_\__{IG}$) in (3a) whereas it does not in (3b). At the segment containing the subcategorizing verb (i.e., *had angered*), native speakers' reading times were shorter in (3a) than in (3b). Following Gibson and Warren (2004), the results were taken to indicate that reactivation of the filler at the intermediate gap site shortens the distance of the filler-gap dependency and consequently facilitates reactivation of the filler at the original gap site. Unlike in native speakers, L2 learners' reading times did not reveal a reading time advantage of (3a) compared to (3b). Marinis et al. concluded that L2 learners do not compute intermediate gaps when processing long-distance filler-gap sentences regardless of whether the learners' L1 involves overt wh-movement or not.

Both Felser and Roberts (2007) and Marinis et al. (2005) show the presence of grammatically-mediated facilitation of filler reactivation for native speakers and the absence of such an effect for late L2 learners who are proficient in L2. This, together with the lack of influence of L1 grammar, constitutes evidence supporting the Shallow Structure Hypothesis. However, the findings are not without alternative explanations. It should be first noted that the stimuli of both studies were quite complex and comprehending them might have been demanding for the learners. Processing the stimuli may have been especially demanding for the learners in the Felser and Roberts' study, in which the participants listened to the auditory stimuli recorded at a normal speed. It is possible that the burden of phonological processing, together with the burden imposed by the dual task, might have prevented the learners from keeping up with the rapidly incoming auditory stimuli and building incremental representations of the input. The position *non-specific* advantage of identical targets therefore may not indicate maintained activation of the filler but may just reflect an advantage of a word heard in the sentence compared to the word that had not been heard in the sentence.

Another aspect of grammar whose role in online L2 processing has been discussed is inflectional morphology. Through a number of offline studies, it is well-known that inflectional morphology in L2 is persistently difficult to acquire and is easily subject to fossilization for late L2 learners (e.g., Lardiere, 1998; White, 2003). The source of difficulty with inflectional morphology has been debated in the literature, with some attributing it to representational deficit, and with others to a problem in retrieving the correct form in real time. More recently, studies use online tasks to examine this issue of source of difficulty with inflectional morphology.

Jiang (2004) asked native speakers of English and proficient Chinese learners of English to read sentences such as (4) in a self-paced reading task to examine whether proficient L2 learners are sensitive to the English plural morpheme.

(4) a. The key to the cabinet was rusty from many years of disuse.

b. The key to the cabinets was rusty from many years of disuse.

Both (4a) and (4b) involve a singular subject (*the key*) and a singular verb (*was*), agreeing in terms of number. In (4b), however, an NP that occurs before the verb has the plural morpheme (*the cabinets*). Previous studies have found that native English speakers' processing of subject-verb number agreement is often influenced by the number feature of the intervening noun, leading to the ‘broken agreement’ effect. Therefore, reading times after the verb were often longer in sentences such as (4b) than in sentences such as (4a) (e.g, Pearlmutter et al., 1999). Jiang took the broken agreement effect as evidence of sensitivity to the plural morpheme and examined whether proficient L2 learners also show this effect in online processing. Results showed that native English speakers took more time to read sentences such as (4b) than sentences such as (4a) at the verb (*was*) and at the immediately following word (*rusty*). L2 learners, however, did not show the broken agreement effect at either word. This finding led to the conclusion that L2 learners are not sensitive to the number morpheme in a comprehension-based online reading task.

Jiang admits that the broken agreement effect does not examine the usual subject-verb agreement processing, but may involve processing strategies unique to native speakers, such as erroneous percolation of the number feature from the local noun rather than from the head noun, to the subject NP. L2 learners’ lack of broken agreement effect, therefore, could suggest that they do not use this processing strategy (and not that they have a problem with the number agreement processing). To address this issue, Jiang therefore conducted a follow-up experiment that tapped directly into subject-verb agreement processing using sentences such as (5).

(5) a. The bridges to the island were about ten miles away.

b. *The bridge to the island were about ten miles away.

Unlike (5a), (5b) involves number disagreement between the subject (*the bridge*) and the verb (*were*). Proper number agreement processing in real time comprehension will cause reading time slow-down at the verb or at the following words in the ungrammatical sentences compared to the grammatical sentences. Results showed that native speakers' reading times were significantly slower in sentences such as (4b) than sentences such as (4a) at the word immediately following the verb (*about*), but no such slow-down was found for the L2 learners. Based on the lack of a 'disagreement effect' for L2 learners, together with the lack of a 'broken agreement effect', Jiang concluded that L2 learners are not sensitive to number agreement in real time comprehension, and further concluded that L2 learners do not have integrated knowledge of morphology.

Studies such as Jiang (2004) suggest that even proficient L2 learners could have problems with sentence-level morphological processing in an L2, consistent with the well-known difficulty with morphology in L2 acquisition. However, there is some evidence that even L2 learners' morphological processing can approximate native speakers' processing with near native-like proficiency. Hopp (2006) investigated how subject-object ambiguities in German are resolved by English and Dutch learners of German, focusing on the role of proficiency. This study included not only advanced learners but near-native learners of German as well. German has base SOV order, but objects can move over the subject. Although the grammatical roles can be overtly signaled by overt case markers (6a-b), the syntactic status of NPs can be locally ambiguous due to syncretistic case morphology (6c-d). Previous research has shown that native speakers of German prefer SOV order to OSV order, and both types of disambiguation (via case at the first NP of the embedded clause (6a-b) and via verbal agreement at the sentence final verb (6c-d)) elicit local reanalysis effects. These processing preferences and the reanalysis effects are determined by the interaction of universal phrase-structure parsing principles and the type of syntactic features that disambiguates order (case vs. verbal agreement).

(6) a. Er denkt, dass **der** Physiker am Freitag **den** Chemiker gegrüsst hat.

He thinks that the_{NOM} physicist on Friday the_{ACC} chemist greeted has

b. Er denkt, dass **den** Physiker am Freitag **der** Chemiker gegrüsst hat.

He thinks that the_{ACC} physicist on Friday the_{NOM} chemist greeted has

c. Sie sagt, dass **die** Baronin am Freitag **die** Bankiers eingeladen **hat**.

She says that the $baroness_{SG}$ on Friday the $bankers_{PL}$ invited has

d. Sie sagt, dass **die** Baronin am Freitag **die** Bankiers eingeladen **haben**.

She says that the $baroness_{SG}$ on Friday the $bankers_{PL}$ invited have

In a self-paced reading study, all L2 learner groups demonstrated preference for SO order. However, learners' proficiency influenced whether they conducted native-like incremental syntactic reanalysis. Although both native speakers and near-native learners showed locally specific slowdowns for OS sentences (for sentences like 6b: at ***den*** *Physiker am Freitag*, and for sentences like 6d: at ***haben***), the advanced learners did not show corresponding slowdowns, suggesting that they haven't converged on incremental native reanalysis patterns yet. This study thus suggests that L2 proficiency plays a role in the depth of morpho-syntactic processing. Although native-like incremental comprehension of inflectional morphology may not be possible even at advanced proficiency, it is possible to develop native-like sensitivity to inflectional morphology in real-time comprehension at a near-native level of proficiency.

Hopp (2010) reported a series of experiments that used materials similar to (6). In a speeded grammaticality judgment task, the study found a difference among the near-native speakers such that at higher speed, performance on case marking dropped to chance levels for the near-natives whose L1 does not instantiate the same grammatical markings as the target language

(English and Dutch) but it did not for those whose L1 has similar grammatical markings (Russian). Moreover, when native speakers conducted speeded grammaticality judgments at different presentation speeds, their performance systematically replicated form-specific difficulties of L2 learners (see McDonald, 2006, for similar findings). Hopp's studies therefore provide important evidence that individual differences in L2 proficiency, learners' L1 and working memory capacity can modulate L2 morpho-syntactic processing.

ERP studies

A considerable portion of L2 processing studies used behavioral measures such as self-paced reading, eye-tracking and cross-modal priming to investigate the way the learners make use of morpho-syntactic information in real-time comprehension. Another set of studies examined neurocognitive mechanisms underlying L2 processing by examining brain responses of L2 learners to linguistic stimuli. One of the most widely used methods is event-related potentials (EPRs), which has a very fine temporal resolution. Since brain activation patterns related to language processing ("ERP components") in the native language are widely studied, in particular with regard to correlates for lexical-semantic processing and morpho-syntactic processing, it is useful for studying L2 processing at these levels. It is known that difficulties in lexical/semantic processing produce an 'N400' signal, a central/posterior bilaterally distributed negativities occurring about 300-500ms after the onset of the stimulus. Difficulties in morpho-syntactic processing elicit other distinct ERP components such as the LAN (left-distributed and/or anterior negativities observed about 100-500ms after the stimulus onset), or P600 (late centro-parietal positivities that start about 500ms after the stimulus onset).

Weber-Fox and Neville (1996), in one of the well-known early ERP studies, examined bilingual speaker's ERPs when they processed semantic and syntactic anomalies, with a goal to investigate whether brain maturation has different effects on the development of distinct subsystems of language. The study tested Chinese-English bilinguals who were exposed to English at different ages (1-3, 4-6, 7-10, 11-13, after 16) and had lived in the US for a minimum of 5 years and attended school or worked during that period. The participants read English sentences such as (7-8) and judged whether they are good English sentences.

(7) semantic/pragmatic control and violation

a. The scientist criticized Max's proof of the theorem.

b. The scientist criticized Max's event of the theorem.

(8) phrase structure control and violation

a. The scientist criticized a proof of the theorem.

b. The scientist criticized Max's of proof the theorem.

All bilingual groups showed N400 in response to violations of semantic expectations such as (7b) compared to control sentences such as (7a), although for the bilinguals who were exposed to English later (between 11-13 and after 16 years of age), the N400 peak latency was slightly longer. However, an age effect occurred when processing sentences such as (8), which involve a phrase structure violation. Although the 1-3, 4-6 and 7-10 groups showed the P600 effect to the phrase structure anomalies such as in (8b), the effect was delayed in the 11-13 group and no P600 effect was found in the >16 group. Therefore, bilinguals showed native-like ERP correlates in response to semantic violations, regardless of age of exposure to the L2, but the ERP patterns in response to syntactic violations showed an age effect, with bilinguals who were exposed to the L2 later showing a delayed ERP response or even no response. The results were taken as evidence in favor of the view that delayed exposure to language impacts development of neural systems relevant for language, and semantics vs. syntax of language are subject to different sensitive periods.

However, there are also ERP studies showing that late L2 processing can be developed such that it can display at least some of the native-like ERP components for processing syntactic anomalies. Tanner, McLaughlin, Herschensohn and Osterhout (2013), for example, tested two groups of native English speakers enrolled in college-level German courses. The first group was novice learners in first-year courses and the other group had more instructions (enrolled in third-year courses). The learners and the native German speakers read sentences that involve incorrect subject-verb agreement and its grammatical counterpart (e.g., *Ich wohne*/**wohnt in Berlin*., "I live/*lives in Berlin.") and judged whether the sentences were good or bad after reading the final word of the sentences. The native speakers produced a P600 effect. The third-year learners also produced a similar P600 effect, suggesting that they detected the subject-verb disagreement in a way similar to native speakers. The first-year learners, however, showed a biphasic N400-P600 response. Follow-up analyses revealed that the biphasic ERP response in the first-year learners was in fact an artifact of averaging over individuals and the first-year learners either showed an N400 dominance or a P600 dominance. The results were taken to indicate that the learners progress through two distinct stages of learning/processing, from for example, the initial stage in which the learners rely on the shallow, good-enough syntactic parse, to the stage in which they conduct deeper grammatical processing like native speakers, but there is variability in the individual's learning speed.

The cross-sectional research design used in Tanner et al. does not guarantee the developmental interpretation. However, the interpretation is supported by another study conducted by the group, which used a longitudinal design (McLaughlin, Tanner, Pitkanen, Frenck-Mestre, Inoue, Valentine & Osterhout, 2010 for a preliminary report). The study tracked English speakers learning French in the first-year university classroom, at three sessions (approximately 4 weeks, 16 weeks and 26 weeks of instruction). ERPs were recorded in response to the subject-verb agreement in French, which also exists in English, and to the number agreement between a definite determiner and noun (*le livre*/*les livres* "the book/the books"), which does not exist in English. For the subject-verb agreement violations, learners showed an N400 effect in the first session, but produced a P600 effect at the last session. In the second session, a subset of the learners primarily showed an N400 response and the others primarily showed a P600 response, suggesting an individual difference in learning speed. For the noun agreement rule, however, the learners' brain responses were not different between the well-formed sentences and those involving disagreement and did not show a change throughout the period of instruction. The results therefore support Tanner et al.'s interpretation that the learners' brain responses can change from a non-native-like to native-like one, even within the first year of instruction. The results also show an effect of L1-L2 similarity, such that the development is more likely to be faster for the rule that is similar to the rule in the learners' L1.

To summarize, previous studies have produced mixed results with regard to adult L2 learners' sensitivity to syntactic information in online sentence processing. There are studies suggesting that L2 learners—even those with advanced proficiency— have difficulty in fully incorporating syntactic information, and this evidence was found in many aspects of processing, including phrase-structure-based preferences in ambiguity resolution, gap processing, sentence-level morphological processing and application of phrase-structure rules. However, there is also growing behavioral and neurological evidence suggesting that full grammatical processing may indeed be possible in L2, given sufficient L2 proficiency, amount of exposure to L2 and cognitive resources, or provided that L1 and L2 are similar in the relevant grammatical aspect. Given that the first group of studies mostly relied on null results, whereas the second group of studies produced positive results, the latter appears to be arguably stronger. Nevertheless, the inconsistency in the literature necessitates further investigation on the nature of syntactic processing in L2.

The series of experiments in this dissertation were conducted in this context. The dissertation specifically investigated whether adult L2 learners are capable of constructing detailed structural representations during online wh-dependency processing, as attested by their sensitivity to a grammatical constraint, called island constraints (Ross, 1967). This is a good test case for the Shallow Structure Hypothesis because sensitivity to island constraints in wh-dependency processing implies that the parser has successfully computed multiple and *hierarchically-organized* syntactic nodes when forming a *non-local* dependency between the filler (i.e., the *wh*-word) and the gap (Clahsen & Felser, 2006c, p. 565). More explanation on what island constraints are and how they have been dealt with in the syntax and processing literature will be provided in the next chapter.

Chapter 3

Island constraints in syntax and L1 and L2 processing

3.1. Island constraints in English and Korean

3.1.1. Island constraints

Languages have syntactic constructions in which an element is displaced from its original position and forms a dependency with a phonologically null element placed in the original position (‘gap’). For example, English *wh*-questions are constructed by moving a *wh*-word to the beginning of the sentence, as shown in (1). In principle, the distance between the extracted element and the gap is unbounded as can be seen in (2), where *which book* crosses three clauses from its original position. It has been discovered, however, that movement is not unconstrained. There are certain syntactic phrases that prevent extraction out of their domains. Those phrases were identified and collectively called ‘islands’ by Ross (1967), and the term ‘island constraints’ has often been used since then. (3a)-(3d) shows examples of some of the better known islands—a complex NP (one that includes a relative clause), a subject, an adjunct, and a *wh*-clause. The islands are marked with < > in the examples.

(1) *Which book* did Bill read *t* ?

(2) *Which book* did Susan say that Tom thinks that Bill read *t* ?

(3) a. **Which student* must you call <the teacher who punished *t* >? (complex NP island)

b. **What* did <the fact that the star remembered *t* > surprise the fans? (subject island)

c. **Who* did you see Mary <before *t* >? (adjunct island)

d. **What* does the journalist wonder <whether the editor would criticize *t* >? (*wh*-island)

The discovery of the island phenomenon has played a significant role in the development of syntactic theory. Although Ross (1967) identified separate sets of island constraints (complex NP constraint, coordinate structure constraint, sentential subject constraint, etc.), the majority of subsequent work has attempted to provide a unified account for various types of islands. In this line of work, effects of various types of islands are derived from a common set of universal constraints in syntax. One of the earliest such attempts is the Subjacency condition (Chomsky, 1973). The account assumes that long-distance movement is cyclic, that is, an element moves successive-cyclically by landing on intermediate COMP positions, if the positions are not lexically filled. Crucially, movement should not cross more than one bounding node in a single step, and bounding nodes of English are noun phrases (NP) or sentences (S) (NP and S correspond to Determiner Phrase (DP) and Inflectional Phrase (IP), respectively, in more recent theories). Therefore, extraction islands are domains that consist of two or more bounding nodes in a single step of movement. Example (4) shows that (3a), an example of violation of a subtype of the complex NP island constraint (an NP modified by a relative clause), crosses three bounding nodes, thus violating Subjacency.

(4) **Which student* must $[_{IP}$ you call $[_{DP}$ the teacher who $[_{IP}$ punished *t*]]]?

Later works revealed that constraints on movement are something more complex than what can be explained by a simple syntactic condition such as Subjacency. For one thing, the list of islands to account for continued to expand as new types of islands were subsequently uncovered (e.g., complex NPs with a relative clause, complex NPs with a complement, sentential subjects, subjects, wh-clauses, adjuncts, coordinate structures, left-branches, VP-adverbs, negatives and other affective operators, quantifiers, etc.). It was also found that islands do not behave uniformly—although some islands (which is called ‘strong’ islands) prevent extraction across-the-board, other islands (which called ‘weak’ islands) do not block extraction of all phrase types (see Szabolcsi & den Dikken, 2007 for a list of strong and weak islands and what types of elements are sensitive or insensitive to weak islands). Moreover, although English was the language that provided data for the initial work, later studies revealed cross-linguistic differences in island effects (see Phillips, to appear, for a summary of cross-linguistic variation). The richness of the island phenomenon made it difficult to subsume all islands under a unified syntactic explanation, and each of the proposed syntactic explanations accounts for only a subset of the islands well (e.g., Chomsky, 1977, 1981, 1986; Cinque, 1991; Huang, 1982; Lasnik & Saito, 1984, 1992; Manzini, 1992; Rizzi, 1990, 2001). Due to the diversity of the constructions that exhibit island effects, attempts have also been made to explain at least parts of the phenomenon by drawing on non-syntactic constraints in grammar (e.g., Szabolcsi & Zwarts, 1993).

For the purpose of the present dissertation, I will assume a syntactic account such as Chomsky (1986) or Lasnik and Saito (1992), which is a modified version of Chomsky (1986). These accounts have been taken as a standard approach for ‘strong’ islands, such as complex NP islands with a relative clause, subject islands or adjunct islands. (5) shows an example of the type of islands examined in the experiments to be reported in this dissertation (the island is marked with < >) .

(5) **Which book* did < the author who wrote *t* > signed the contract with the publisher?

In (5), the *wh*-word has been extracted out of a relative clause modifying the subject NP, which degrades the grammaticality of the sentence. The bracketed phrase is at least a ‘double’ island, in the sense that it is the subject of the sentence (a subject island) and it also involves a relative clause modification (a ‘relative clause’ island, which is a subtype of the complex NP island). In fact, the relative clause is an adjunct of the subject NP, which can also be categorized as an adjunct island.

In Chomsky (1986)/Lasnik and Saito (1992), movement should obey subjacency and is constrained by ‘barriers’. The definition of barriers and subjacency in Lasnik and Saito’s (1992) framework is given in (6) and (7).

(6) Barrier

γ is a *barrier* for β if

a. γ is a maximal projection

b. γ is not L-marked (θ-governed by a lexical category), and

c. γ dominates β

(7) Subjacency

β is *subjacent* to α if for every γ, γ a barrier for β, the maximal projection immediately dominating γ dominates α. (Lasnik & Saito, 1992, p. 87)

Chomsky (1986) presents a different (and more complicated) definition of barriers and subjacency, but both frameworks share the most crucial insight incorporated in the definitions of barriers—that non-complements block extraction but complements don’t (Huang’s (1982)

Condition on Extraction Domains). This is shown in (6b), which states that barriers are those that are not lexically theta-marked. The Chomsky/Lasnik and Saito framework thus provides a straightforward account of strong islands such as relative clauses, subjects and adjuncts since none of these phrases are complements. (8), which is taken from Lasnik and Saito (1992, p. 88) and slightly adapted, illustrates how their system blocks extraction out of a relative clause island.

(8) *$Where_2$ did you see the $book_1$ $[_{CP}$ $which_1$ $[_{IP}$ John put t_1 t_2]]?

The embedded IP is a barrier for *t*2, since it is a maximal projection (6a), it is not L-marked (6b), and it dominates *t*2 (6c). Since *where* is not dominated by the embedded CP, the maximal projection immediately dominating IP, Subjacency is violated. Subjacency is also violated in relation to the embedded CP, since the embedded CP is not L-marked, but *where* is not dominated by the immediately dominating maximal projection (the NP). The doubly violated Subjacency yields ungrammaticality in (8).

In this dissertation I examine sensitivity to the subject/relative clause island (as in (5)) by Korean L2 learners of English. This type of island produces a strong perception of ungrammaticality in English (the target language), and can be explained by accounts such as Chomsky/Lasnik and Saito. In Korean (the learners' L1), however, applying the Chomsky/Lasnik and Saito-style grammar is less straightforward as will be discussed in the next section.

3.1.2. Island constraints in Korean

Unlike in English, Korean *wh*-questions do not involve overt movement of *wh*-words. Sentences (9a-b) show an example of the declarative sentence and its interrogative counterpart. The scope of the *wh*-word in the interrogative sentence is marked by a question particle at the end of the sentence and the *wh*-word stays *in-situ*, making the word order of declarative sentences and interrogative sentences identical. The language, however, does have constructions in which an element occurs in a non-canonical position, such as scrambling (10) or relativization (11).

(9) a. Chelsu-nun ku chayk-ul sass-ta. (declarative sentence)

Chelsu-top the book-acc bought-dec

'Chelsu bought the book.'

b. Chelsu-nun *etten chayk-ul* sass-ni? (interrogative sentence)

Chelsu-top which book-acc bought-Q

'Which book did Chelsu buy?'

(10) a. Chelsu-nun [Yenghuy-ka ku chayk-ul sassta-ko] sayngkakhanta. (canonical order)

Chelsu-top Yenghuy-nom the book-acc bought-comp thinks

b. *ku chayk-ul* Chelsu-nun [Yenghuy-ka sassta-ko] sayngkakhanta. (scrambled order)

the book-acc Chelsu-top Yeonghuy-nom bought-comp thinks

'Chelsu thinks that Yenghuy bought the book.'

(11) a. Yenghuy-ka ku chayk-ul sassta.

Yenghuy-nom the book-acc bought

'Yenghuy bought the book.'

b. [Yenghuy-ka san] *ku chayk* (relativization)

Yenghuy-nom buy-adn the book

'the book that Yeonghuy bought'

Notably, Korean exhibits island effects less widely than English. First, in *in-situ wh*-questions, *wh*-words that are placed within islands generally do not degrade grammaticality. It is not that Korean completely lacks island effects in *in-situ wh*-questions: for example, it has been reported that the *wh*-word cannot be placed within a *wh*-island (Yoon, 2006), and an adjunct phrase such as '*why*' cannot be the *wh*-word in many islands. However, the lack of island effects in *wh*-questions is observed for most types of islands and for most types of *wh*-words as in (12) and (13). The examples show that no relative clause island effect or adjunct island effect is found in Korean *wh*-questions. Sentences in (14a-b) show that the same is true for the particular island configuration tested in the present dissertation. As in (14a), which is a repetition of (5), *wh*-question formation out of the relative clause/subject island is not allowed in English. However, the Korean translation in (14b) is perfectly grammatical. On the assumption that *in-situ wh*-words move covertly to Comp at Logical Form (LF) (e.g., Huang, 1982; Lasnik and Saito, 1984), the lack of island effects in *wh*-questions is unexpected if constraints such as those proposed by Chomsky/Lasnik-Saito operate in Korean.

(12) Chelsu-nun <*nwu-ka* ssu-n chayk-ul> sass-ni?

Chelsu-top who-nom write-and book-acc bought-Q

'*Who did Chelsu buy the book that was written by *t* ?'

(complex NP (relative-clause) island)

(13) Chelsu-nun <*nwu-ka* nucess-ki ttaymwuney> hwa-ka nass-ni?

Chelsu-top who-nom be late-comp because anger-nom come-Q

'*Who did Chelsu get angry because *t* was late?'

(adjunct island)

(14) a. **Which book* did < the author who wrote *t* > signed the contract with the publisher?

b. <*etten chayk-ul* ssu-n cakka-ka> ku chwulphansa-wa kyeyyakhayss-ni?

which book-acc write-adn author-nom the publisher-with signed the contract-Q

'*Which book did the author who wrote *t* signed the contract with the publisher?'

Island effects are less widely observed in Korean than in English even in constructions involving overt movement. For example, scrambling is permitted out of adjunct conditional clauses (Yoshida, 2006) as in the following example, whereas movement out of the same clause is not allowed in English.

(15) *etten chinkwu-lul* Chelsu-nun <manyak Yenghuy-ka *t* chotayha-myen> wul-kka?

which friend-acc Chelsu-top cond-adv Yenghuy-nom invite-cond will cry-Q

'*Which friend will Chelsu cry if Yenghuy invites t ?'

One finding that is more closely relevant to the island construction tested in the present dissertation is that Korean has so-called 'double relative clauses'. As exemplified in (16), the clause involves relativization out of a relative clause, which is an island. Although the movement crosses a relative clause island, (16) is perfectly grammatical in Korean, unlike its English translation.

(16) < $\_\_\__i$	$\_\_\__j$	cohaha-nun	kangaci$_j$-ka >	cwuk-un	ai$_i$
		like-adn	dog-nom	die-adn	kid

'*the kid$_i$ who the dog $_j$ which t_i liked t_j died' (Han & Kim, 2004, p. 316)

Considering that there is limited evidence of island effects in Korean, a possible approach to this language would be to assume that island constraints do not operate in it. This has been in fact the assumption of many previous studies investigating how L2 learners of English whose L1 is Korean (or other East Asian *wh-in-situ* languages such as Chinese and Japanese) acquire island constraints in English (e.g., Hawkins & Hattori, 2006; Johnson & Newport, 1991; Schachter, 1990; White & Juffs, 1998).

However, there have been many attempts to maintain the universality of island constraints by providing alternative explanations for the lack of island effects in these languages. With regard to the no island effect in *wh-in-situ* constructions (illustrated in (12)-(14)), Huang (1982) proposed that Subjacency applies to overt movement, but it does not restrict LF movement. In another account, Choe (1987) argued that what undergoes LF *wh*-movement is not the *wh*-word alone. Instead, the whole island containing the wh-word moves, consequently avoiding Subjacency violation. It was also suggested that the *in-situ wh*-words are not licensed through movement at LF but through a different mechanism such as selective binding (Aoun & Li, 1993; Tsai, 1994). As for the lack of island effect in double relative clauses, Han and Kim

(2004) proposed a different source sentence for examples such as (16), based on the fact that Korean has double nominative constructions such as ‘the kid-NOM dog-NOM died.’ (meaning ‘As for that kid, the dog died.’). According to their analysis, (16) is actually not ‘double relative clauses’, and movement does not cross a relative clause boundary. What these accounts commonly suggest is that ‘apparent’ violation of island constraints in languages such as Korean, Chinese and Japanese does not imply that these languages are not constrained by the universal syntactic constraints on movement.

Summing up, Korean shows sensitivity to islands in certain cases, but there are much more limited island effects in Korean than in English. Despite this, continued attempts have been made in the literature to maintain the assumption that island constraints apply universally, including in Korean. Under this assumption, the (general) lack of island effects in Korean is due to a surface-level variation and not due to a deeper-level, parametric variation. However, it should be pointed out that even though Korean might be different from English only at a surface-level, it is still possible that the surface-level differences might potentially have an interference effect when Korean L2 learners of English apply island constraints in English, especially in real-time processing. Although the studies included in this dissertation do not directly address the issue of L1 transfer, the L1-L2 differences may potentially influence sensitivity to island constraints in English filler-gap processing by Korean learners.

3.1.3. Island constraints as constraints on processing resources

As reviewed above, the dominant approach to island effects has been to assume that the effects are a consequence of a grammar (a set of syntactic constraints) represented in language. However, it should be mentioned that there have also been attempts to explain island effects by drawing on non-grammatical principles. Among these alternative perspectives, the most well-known group of work accounts for island effects by using independently-motivated constraints on language processing (Deane 1991; Hofmeister & Sag, 2010; Pritchett 1991; Kluender & Kutas 1993) and is sometimes called 'processing-based' accounts or 'resource-based' accounts. According to this approach, there are no formal grammatical constraints that prohibit extraction out of islands; instead island effects occur as epiphenomena of processing difficulty. In other words, extraction out of islands is not actually ungrammatical, but results in a *perception* of ungrammaticality. A merit of this approach is that it can simplify grammar by eliminating complex island constraints from it. Moreover, if the cognitive constraints used to derive island effects are indeed independently motivated (i.e., necessary to explain the language processing behavior anyway), then the approach would be superior to the grammar-based accounts of islands.

Work by Kluender (Kluender, 1998; Kluender & Kutas, 1993) and Hofmeister and Sag (2010) presents a fairly elaborated account for the wh-island and the complex NP island. In this account, processing long-distance *wh*-dependency across the *wh*-island or the complex NP island involves multiple linguistic operations, such as holding a filler in working memory while searching for the gap, and referential processing at the left boundary of the islands. Each of the processes adds a burden to the working memory capacity, but is manageable if occurring alone. However, concurrent occurrence of the two processes, for example forming a *wh*-dependency across the islands, taxes working memory, and the difficulty of processing increases above the threshold of grammaticality. Although the *wh*-island and the complex NP island are the prime examples of this approach, processing-based accounts of other islands have also been proposed (e.g., Kluender, 2004, for an account of the subject island).

Another point to mention is that there has also been an intermediate position between the grammar-based accounts and the processing-based accounts. Under this approach, the island effect originated diachronically from processing difficulty, but has been grammaticized (Fodor 1978, 1983, Berwick & Weinberg, 1984, Hawkins, 1999). Therefore, from the synchronic point of view, island constraints are represented as grammatical knowledge.

The different views on the nature of island constraints (grammar-based vs. processing-based accounts) have important theoretical ramifications in L2 acquisition and processing research. Island constraints have been traditionally considered as grammatical (UG) constraints in L2 research, and sensitivity to island constraints by L2 learners has been interpreted as evidence for UG accessibility in L2 acquisition (e.g., Johnson & Newport, 1991; Schachter, 1991; White & Genesee, 1996). Under this view, sensitivity to island constraints in real time processing could also be interpreted as evidence for deep syntactic processing as in this dissertation. On the other hand, if island constraints are not (synchronically) encoded in grammar and are pure by-products of processing difficulty, sensitivity to islands by L2 learners will instead provide insights on how L2 processing is influenced by processing-related factors. As we will see later, the overall results of this dissertation are more compatible with the view that island constraints are explicitly encoded in grammar.

The next section reviews the studies that have investigated how native speakers apply island constraints in online processing.

3.2. Parsing strategy and island constraints in L1 filler-gap processing

As illustrated in (1), '*Which book* did Bill read *t*', sentences such as English *wh*-questions include a displaced *wh*-word. In this kind of sentences, the displaced word (*which book* in (1)) has been called the ‘filler’ and the empty category (*t* in (1)) in the original position has been called the ‘gap’. In order to interpret the filler-gap sentences, the filler needs to be stored in working memory until it finds the gap, where it is assigned a thematic role. Since the process can be demanding, especially for sentences that involve long-distance movement such as (2), (*Which book* did Susan say that Tom thinks that Bill read *t*), the way filler-gap sentences are processed has been treated as an interesting topic in the psycholinguistic literature.

One of the widely accepted findings in the filler-gap processing literature is that parsing filler-gap sentences is based on a strategy, often called the “active filler strategy” (e.g., Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Clifton & Frazier, 1989; Frazier, 1987; Garnsey et al., 1989; Lee, 2004; Pickering & Traxler, 2003; Stowe, 1986; Sussman & Sedivy, 2003; Traxler & Pickering, 1996). Once the parser identifies a filler, it must find the gap for the filler in order to form a dependency between them. Although, in principle, the parser can wait until it encounters clear bottom-up evidence for the gap position (i.e., a missing argument), it instead actively predicts where the gap might be and forms a dependency as soon as it finds the earliest possible gap position. This strategy leads to the correct parse in many cases, as seen in example (10), but it does not guarantee the correct analysis. For example, in (11), the filler *which book* is initially interpreted at t_1, that is, as the object of *write* (*Which book did the author write?*), because this is the first possible gap position. However, after encountering the word *while*, it becomes clear that *about* has a missing argument and the filler actually needs to form a dependency with t_2. The parser is thus forced to revise its initial interpretation.

(17) *Which book* did the author write *t* ?

(18) *Which book* did the author write (t_1) passionately about t_2 while he was traveling?

The reason why the parser forms a filler-gap dependency as soon as possible has been attributed to a more general processing principle that requires the parser to complete grammatical dependencies as soon as possible (de Vincenzi, 1991; Frazier, 1987; Pritchett, 1992), or the need to reduce the cost of retaining the filler in memory (Gibson, 1998).

Given the parser's tendency to use the active filler strategy, previous research has asked whether the parser actively creates a gap even if active gap creation violates grammatical constraints, that is, island constraints. Specifically, studies have examined whether the filler can be associated with the closest potential gap even if the gap is within an island. If island constraints immediately constrain real-time structure building, the usual active gap search strategy should be suspended within an island to prevent illegal gap creation. Most studies that examined this issue have focused on filler-gap processing in English.

Event-related potentials (ERP) studies examining this issue in English found that island constraints operate immediately in filler-gap processing. McKinnon and Osterhout (1996), for example, analyzed the ERPs of participants who read sentences like (19a)-(19b).

(19) a. *I wonder which of his staff members the candidate was annoyed when his son was questioned by.

b. I wonder whether the candidate was annoyed when his son was questioned by his staff member.

(19a) involves an adjunct island violation: the *wh*-phrase has been extracted out of an adjunct clause (i.e., *when his son was questioned by*), whereas (19b) does not involve such a violation. McKinnon and Osterhout found a large, widely distributed P600 response (i.e., a response that is typically associated with syntactic anomalies) immediately after *when*. This result, together with findings of other ERP studies that are compatible with these results (Neville et al., 1991; Kluender & Kutas, 1993) and speed-accuracy tradeoffs (McElree & Griffith, 1998), suggests that island boundaries (i.e., the first indication of island violation) are detected. Although these studies suggest sensitivity to island boundaries, they do not show whether or not the parser actually suspends the formation of filler-gap dependencies within islands.

Two studies provide more direct evidence that active gap creation is blocked by island constraints. Stowe's (1986) self-paced reading study used the 'filled-gap effect' to investigate whether the parser suspends the active gap search within a subject island. A filled-gap effect refers to an increased reading time in the position where the parser predicts a gap, but encounters an overt NP instead. Because the overt NP signals that the parser's expectation of a gap is not correct, reading time increases locally. Stowe used sentences like (20).

(20) a. The teacher asked what the team laughed about Greg's older brother fumbling.

b. The teacher asked if the team laughed about Greg's older brother fumbling the ball.

c. The teacher asked what the silly story about Greg's older brother was supposed to mean.

d. The teacher asked if the silly story about Greg's older brother was supposed to mean anything.

Sentences (20a) and (20c) include the fronted *wh*-word *what*. In both sentences, the first potential gap site follows the preposition *about*, but it is occupied by the overt NP *Greg's older brother*. In (20c), this potential gap site is inside a subject island (the embedded subject, *the silly story about Greg's older brother*), whereas in (20a), it is not inside an island. Sentences (20b) and (20d), which are not filler-gap sentences, were used as control sentences. Stowe found a filled-gap effect at *Greg's* in (20a), as compared with (20b), indicating that the parser was 'surprised' in (20a), when *Greg's* was found in the position where it expected a gap. However, a similar effect was not found in (20c)-(20d), suggesting that the parser did not expect a gap to exist within an island.

Traxler and Pickering's (1996) eye-tracking results also show that the normal active gap search is suspended within an island. This study used a plausibility manipulation paradigm as illustrated in (21a)-(21d).

(21) a. We like the book that the author wrote unceasingly and with great dedication about while waiting for a contract.

b. We like the city that the author wrote unceasingly and with great dedication about while waiting for a contract.

c. We like the book that the author who wrote unceasingly and with great dedication saw while waiting for a contract.

d. We like the city that the author who wrote unceasingly and with great dedication saw while waiting for a contract.

In (21c)-(21d), but not in (21a)-(21b), the dependency between the head noun of the relative clause (*the book* and *the city*) and the first potential gap (after *write*) spans an island (*the author who wrote unceasingly and with great dedication*). The island can be considered as a 'double' island because it is both the subject of the embedded clause and includes a relative clause. The plausibility of the relative-clause head noun as object of the first potential subcategorizing verb was manipulated within the non-island and island conditions (i.e., *the book*, but not *the city*, is a plausible object of *write*), as a diagnostic for the dependency between these words. Traxler and Pickering found a plausibility effect at the first potential gap site in the non-island conditions: the reading times were longer when the filler was not a plausible object, as in (21b), than when it was a plausible object, as in (21a), which was taken as evidence that the filler was interpreted as the object of the first verb. However, such a plausibility effect was not found in the island conditions. These results suggest that the parser establishes a local filler-gap dependency only when there is no intervening island.

In sum, the majority of the existing studies suggest that native speakers do not construct a gap that violates island constraints in real-time processing. Although there exist a small number of studies indicating that the parser forms a filler-gap dependency within an island (Clifton & Frazier 1989; Freedman & Forster, 1985; Pickering et al., 1994), the standard view has been that filler-gap processing immediately obeys island constraints (see Phillips, 2006, for a discussion on the possible reasons for the insensitivity to islands in these studies).

Immediate sensitivity to island constraints was recently taken as evidence for the view that espouses a tight relationship between grammar and parser—i.e., real-time sentence comprehension in a native language makes use of accurate and full-fledged grammatical representation (Phillips, in press; Phillips, Wagers & Lau, 2011). Based on the previous finding that L1 filler-gap processing is constrained by island constraints, this dissertation sought to investigate whether L2 filler-gap processing is also constrained by the constraints. If L2 processing turns out to be sensitive to island constraints, this would suggest that L2 processing is as grammatically precise as L1 processing, and would in turn suggest that it is done over detailed syntactic representations. Before presenting the methods and results of the experiments, we turn to a review of relevant literature in L2 research below.

3.3. Island constraints in L2 acquisition and processing

One of the central theoretical debates in L2 acquisition research centers on whether adult L2 learners have access to Universal Grammar (UG) (Chomsky, 1965), the innate linguistic knowledge that is thought to be involved when children acquire their native language. The innate knowledge is about the kinds of grammar that characterize human language and the kinds of crosslinguistic variation in grammar. UG limits the implicit hypotheses children consider for given language input, enabling them to develop uniform underlying grammatical competence rapidly. At issue with regard to L2 acquisition is whether this innate linguistic knowledge constrains acquisition of a second language by adult learners. That is, whether UG is accessible when the second language input is given after the period of time that is often considered as maturationally scheduled for language acquisition (e.g., Bley-Vroman, 1990, 2009, Schwartz & Sprouse, 1996).

As a UG principle, island constraints (usually called the 'Subjacency Principle' in the L2 literature) have received significant attention in this theoretical context. A number of experiments have been conducted in order to examine L2 learners' sensitivity to island constraints in L2 (e.g., Bley-Vroman, Felix, & loup, 1988; Hawkins & Chan, 1997; Hawkins & Hattori, 2006; Johnson & Newport, 1991; Martohardjono & Gair, 1993; Schachter, 1990; White, 1992; White & Genesee, 1996; White & Juffs, 1998). The studies have produced results that were interpreted as support for various theoretical perspectives with regard to the UG accessibility issue, ranging from the full UG access view to no UG access view. One point that is worth noting about these studies in relation to the current dissertation is that learners whose L1 is a *wh*-in situ language, such as Chinese, Japanese and Korean, were generally found to be insensitive to violations of island constraints even at advanced L2 proficiency level (e.g., Hawkins & Chan, 1997; Hawkins & Hattori, 2006; Johnson & Newport, 1991; Schachter, 1990), although there is evidence to the contrary as well (e.g., White & Juffs, 1998). Poor performance of this learner group was taken by many researchers as evidence against full UG access in L2 acquisition, leading them to propose that there is a maturational decline in access to island constraints by L2 learners (Johnson & Newport, 1991) or island constraints are not accessible to adult L2 learners if the constraints have not been triggered during L1 acquisition (Schachter, 1990).

The existing studies on acquisition of island constraints by L2 learners are wide in scope in that they have examined many types of islands and have tested learners with various L1 backgrounds. However, the studies traditionally had a limited scope from the methodological point of view. Most of them relied on offline tasks such as one that asks participants to read/listen to an entire sentence and judge grammaticality, which provide information primarily on how island constraints influence participants' final interpretation/judgment. Using available psycholinguistic techniques, however, it is possible to examine the role of island constraints at different processing stages before reaching the final interpretation/judgment. Studying the online processing patterns potentially provides additional information on the state of knowledge of island constraints by L2 learners, which is one of the contributions of the present study to the literature of L2 acquisition of island constraints.

As reviewed above, the phenomenon of active gap search was used as a diagnostic tool in investigation of the influence of island constraints in L1 processing. That is, whether or not the parser suspends active gap search in an island domain reveals whether it is sensitive to island constraints. In order to utilize this research method in L2 processing, a prerequisite is that L2 learners also use the active gap search strategy to process filler-gap sentences. So far, L2 processing research has indeed shown that L2 learners actively predict the gap at the earliest possible location, as do native speakers (Aldwayan, Fiorentino & Gabriele, 2010; Juffs & Harrington, 1995; Schachter & Yip, 1990; Williams, Möbius & Kim, 2001), which makes it feasible to adopt the research design of L1 processing studies in L2 research.

Among these studies, Williams et al. (2001) is worth closer examination because they used the plausibility manipulation paradigm as in the experiments in this dissertation. Chinese, German and Korean L2 learners of English read sentences such as (23) one word at a time and indicated as soon as sentences did not make sense (‘stop-making-sense’ task).

(23a) Which girl did the man push the bike into late last night?

(23b) Which river did the man push the bike into late last night?

The sentences were designed to investigate active gap search as well as sensitivity to plausibility in L2 processing. As the first available subcategorizer after the filler, the verb was predicted to induce gap postulation if L2 learners use the active gap search strategy. However, the post-verbal object NP signaled that the posited gap is not licit and must be cancelled. The gap must ultimately be posited after the preposition, as its object. The plausibility of the filler-gap association was manipulated at the verb such that it was plausible in (23a) and implausible in (23b). Using the manipulation, the study examined whether L2 learners evaluate the plausibility of the initially posited wh-dependency rapidly, and whether they use this plausibility information for reanalysis. The learners' implausibility detection rate was higher in (23b) than (23a) at the verb and the following determiner (e.g., *push the*), suggesting that they identified a gap upon reading the verb and rapidly evaluated the plausibility of the filler-gap dependency. Moreover, the fact that the implausibility detection rate was higher in (23a) than (23b) from the postverbal noun (e.g., *bike*), appears to indicates that the plausibility of the initially formed wh-dependency affected the learners' recovery from the misanalysis (i.e., plausible initial parse is more difficult to abandon). Williams et al.'s study showed that the Korean learners of English, which is the target population of this dissertation, actively form a wh-dependency at the earliest possible chance and evaluate the plausibility of the dependency, at least when the task facilitates incremental semantic processing. Building on this earlier finding, this dissertation examined whether Korean learners of English actively search for a gap and evaluate plausibility within an island construction, as a way to test sensitivity to island constraints of this group.

Interest in the role of island constraints in real time processing arose recently among L2 researchers, and several initial sets of processing studies were conducted independently from each other, including the first experiment of this dissertation. Since some of these studies are available to read by the time this dissertation is completed, they will be discussed in this section. Aldwayan et al. (2010) investigated whether advanced Najdi Arabic learners of English produce a filled-gap effect within an island using the self-paced reading task.

(24a) My sister wondered if the boring comments about John's used car were intended to entertain the group.

(24b) My sister wondered who the boring comments about John's used car were intended to entertain ___.

The learners encountered a filler while reading (24b) (i.e., *who*), whereas they did not while reading (24a). If they interpret the filler as the object of the first subcategorizer (i.e., *about*) in (24b), the reading time will slow down at the overt object of this subcategorizer (i.e., *John's used car*) compared to the same phrase in (24a). However, the learners did not produce such filled-gap effects, suggesting that they did not search for a gap within the complex subject of the embedded clause (i.e., *the boring comments about John's used car*), which is an extraction island.

A different group of studies (Cunnings, Batterham, Felser & Clahsen, 2010; Omaki & Schulz, 2011) manipulated plausibility as in (25) to examine the role of island constraints in L2 English processing (example from Omaki & Schulz, 2011).

(25a) The city that the author wrote regularly about was named for an explorer.

(25b) The book that the author wrote regularly about was named for an explorer.

(25c) The city that the author who wrote regularly saw was named for an explorer.

(25d) The book that the author who wrote regularly saw was named for an explorer.

In (25c-d) but not in (25a-b), the relative clause head noun *city/book* and the first subcategorizer *wrote* spanned across a relative clause (*the author who wrote regularly*), which is an extraction island. Plausibility manipulation between the filler and the first subcategorizer in the non-island and island conditions made it possible to test whether the gap is created at the first subcategorizer within a relative clause. Testing the German and Chinese-speaking learners (Cunnings et al., 2010) and Spanish-speaking learners (Omaki & Schulz, 2011), both studies found that a plausibility mismatch effect occurred in (25a-b) but not in (25c-d), indicating that their learners were sensitive to island constraints.

One thing to mention about Cunnings et al. (2010) and Omaki and Schulz (2011) is that although the two studies used similar research designs and found compatible results, they came to different conclusions. Assuming that island constraints are syntactic constraints in grammar as in most previous L2 studies, Omaki and Schulz (2011) took the results to indicate that L2 learners are able to apply island constraints rapidly in real time processing and concluded that the learners are able to build syntactic representations with grammatical precision. Cunnings et al. (2010), however, adopted the processing-based account of islands (e.g., Hofmeister & Sag, 2010; Kluender & Kutas, 1993), which claims that the island phenomenon is non-grammatical in nature and it occurs due to difficulty in retrieving the filler while processing certain resource-demanding constructions. According to Cunnings et al., sensitivity to islands by L2 learners is caused by the learners' limited processing capacity and is irrelevant to the issue of how precise L2 grammatical processing is. These two studies constitute an example showing that studies on the role of a certain construction in L2 processing can lead to different theoretical interpretations depending on the nature of the given construction. In future studies examining the role of island constraints in L2 processing, an effort also needs to be made to empirically test whether the island phenomenon comes from grammatical constraints or processing constraints. As one way to reveal the source of island effects in the native speaker group and in the L2 learner group, this dissertation explored the relationship between sensitivity to islands and individual working memory capacity (Experiment 2 and 3; cf, Sprouse et al., 2012). As we will see, the overall results are more compatible with the grammar-based account of island constraints than with the processing-based account.

The dissertation includes three experiments, all testing Korean learners of English, which is a learner group that has not yet been examined in L2 island processing research. Experiment 1 investigated learners' sensitivity to island constraints using the stop-making-sense task. Experiment 2 used the eye-tracking technique to investigate sensitivity to island constraints in an ecologically more valid reading situation, and examined the role of individual working memory capacity. Experiment 3 examined learners' sensitivity to the difference between two types of islands, which requires an even higher level of grammatical precision. The methods and results of each experiment will be presented in the following three chapters.

Chapter 4

Experiment 1

The primary goal of the experiment reported in this chapter was to investigate whether processing of *wh*-dependency in L2 English by Korean learners of English is constrained by island constraints. Experiment 1 examined the effect of one type of island—a relative clause embedded within a complex subject, which was first explored in a native language processing study by Traxler and Pickering (1996). An offline grammaticality judgment task and an online stop-making-sense task (Boland et al., 1995) were conducted to reveal whether the learners have grammatical knowledge of island constraints and how they use the knowledge in online sentence comprehension. In addition to this primary goal, Experiment 1 also describes the way the learners search for a gap and reanalyze an incorrectly formed wh-dependency.

4.1. Method

4.1.1. Participants

The L2 learners of this study were 31 Korean learners of English (age 20-38, mean = 26.8) studying at University of Illinois at Urbana-Champaign as an undergraduate or a graduate student. Twenty-four native speakers of English (age 19-34, mean = 21) studying at the same university participated as the control group. The L2 learners filled out a language background questionnaire at the beginning of the experimental session. The questionnaire included questions asking the participants' age at the time of testing, the age at which L2 learners began learning English in Korea, the duration of their stay in the US, and the amount of English use in a day.

Table 4.1 summarizes the learners' language background. All learners reported that they had been living in Korea before they arrived in the US. In Korea, daily communication is almost exclusively carried out in Korean, and exposure to meaningful English use is usually confined to special venues such English classes and some university classes delivered in English. Although the possibility cannot be excluded that some of the learners had individually created opportunities to use English for a significant amount of time, it appears to be reasonable to assume that the learners as a group had been immersed in an English-speaking environment only after they arrived in the US. Since the learners' age at the time of arriving in the US was 15-34 (mean=23.1), at least most of the learners of the experiment can be characterized as adult (post-puberty) learners, whose genuine exposure to English began past the so-called critical period of language learning (Lenneberg, 1967). Table 4.1 also shows that there was quite a large amount of variation in the learners' duration of stay in the US and the self-reported daily English use.

Table 4.1. L2 learners' language background information

		age of first	age of first	duration of	English use
Korean speakers	range	7-14	15-34	less than 1-8	5-100
	mean	11.1	23.1	3.6	40.8
	SD	1.9	5.1	2.4	26.8

A written cloze test, presented in Appendix B, was administered to measure the learners' general English proficiency. The passage of the test was adapted from an English teaching textbook (O'Neill, Cornelius & Washburn, 1991) and the reliability of this test is high (Cronbach alpha=.817 according to Ionin & Montrul, 2010). The test included 40 blanks, each with three choices, and each correct answer was counted as 1 point. Out of the maximum score of 40, the learners' mean score was 33.6 (range=28-37, SD=2.4). Although the test was not administered to the native speakers in the present experiment, another group of native English speakers who participated in a previous study (Ionin, Baek, Kim, Ko & Wexler, 2012) scored 38.1 (range=36-39). The English proficiency of the L2 learners of Experiment 1, therefore, was lower than those of native English speakers, but was fairly high.

4.1.2. Materials

The main test consisted of an online and an offline task. The online task was the stop-making-sense task, which had been used to study online filler-gap processing in both native language (e.g., Boland et al., 1995) and second language (Williams, 2006; Williams et al., 2001). The offline grammaticality judgment task was administered to test whether the L2 learners had acquired the relevant island constraints in English because having grammatical knowledge is a prerequisite for using it in online comprehension. In case the learners do not show evidence of using island constraints in online processing, the offline task was expected to reveal whether the online insensitivity to island constraints is due to the learners' lack of grammatical knowledge, or due to failure to use the already acquired grammatical knowledge in real time comprehension.

The stop-making-sense task included 20 sets of experimental sentences. Four conditions were created by manipulating two factors—Island and Plausibility—as in Traxler and Pickering (1996). (1a)-(1d) shows an example experimental item. The groups of words that were presented together to the participants ('regions', which is abbreviated as 'R' in (1)) were indicated by the slash bars ("/"). Participants did not see the slash bars or the region marking at the time of testing.

(1) a. Non-island/Plausible

I wonder $_{R1}$ / which book $_{R2}$ / the author $_{R3}$ / wrote passionately $_{R4}$ / about $_{R5}$ / while he $_{R6}$/ was traveling $_{R7}$.

b. Non-island/Implausible

I wonder $_{R1}$ / which city $_{R2}$ / the author $_{R3}$ / wrote passionately $_{R4}$ / about $_{R5}$ / while he $_{R6}$ / was traveling $_{R7}$.

c. Island/Plausible

I wonder $_{R1}$ / which book $_{R2}$ / the author $_{R3}$ / who wrote passionately $_{R4}$ / saw $_{R5}$ / while he $_{R6}$ / was traveling $_{R7}$.

d. Island/Implausible

I wonder $_{R1}$ / which city $_{R2}$ / the author $_{R3}$ / who wrote passionately $_{R4}$ / saw $_{R5}$ / while he $_{R6}$ / was traveling $_{R7}$.

The experimental sentences included a filler (e.g., *which book/city*) and a possible gap location after the first embedded verb (e.g., *wrote*), which is earlier than the actual gap position (e.g., after *about/saw*). The earlier possible gap location was not inside an island in (1a)-(1b) whereas it was inside an island *the author who wrote passionately* in (1c)-(1d). This phrase functions as an island because it is a complex noun phrase including a relative clause (complex NP island or relative clause island), and at the same time it serves as the subject of the embedded clause (subject island). The plausibility of the fillers as direct object of the first embedded verb was manipulated within each Island condition (e.g., *which book,* but not *which city*, was a plausible object of *wrote*). The plausibility manipulation resulted in an interpretation of the sentence that was temporarily plausible or implausible at the earliest gap position, but only for sentences that did not contain an island; for sentences that contained an island, a gap should not be posited after the verb, and thus there should not be an effect of plausibility. In other words, the plausibility manipulation should reveal whether a gap is posited at the earliest possible position, and the island manipulation should determine whether the relative-clause island prevents the early gap search. All the sentences were globally plausible at the actual gap location. Optionally transitive verbs were used as the first embedded verb (e.g., *write*, *sing*, *wash*) in order to render both the temporary and ultimate filler-gap associations grammatical. A complete list of the experimental sentences of the stop-making-sense task can be found in the Appendix A.

The participants were randomly assigned to one of four presentation lists. Each participant read only one sentence in an experimental item and an equal number of sentences across the four conditions. The experimental items were interspersed with 44 filler sentences, which were indirect *wh*-questions. One half of the filler items were globally plausible and the other half were globally implausible.

To ensure that the plausibility of the filler as object of the first embedded verb was different in the plausible and implausible conditions, a plausibility test was conducted with additional 10 native speakers of English. Interrogative sentences such as those in (2a)-(2f) were constructed from each set of experimental sentences. They tested the plausibility of the filler-gap association at the temporary gap position in (2a)-(2b), at the actual gap position of the Non-island condition in (2c)-(2d), and at the actual gap position of the Island condition in (2e)-(2f).

(2) a. Which book did the author write?

b. Which city did the author write?

c. Which book did the author write about?

d. Which city did the author write about?

e. Which book did the author see?

f. Which city did the author see?

The participants saw 120 sentences (6 conditions x 20 items) in random order and rated their plausibility on a 7-point scale, with 1 being ‘very implausible’ and 7 being ‘very plausible’. If plausibility was manipulated as intended, condition b should be judged as less plausible than condition a. However, the plausibility ratings were expected to be comparable at the actual gap positions (between conditions c and d and between conditions e and f). The results showed that condition b had a lower mean plausibility rating (mean=2.82, SD=1.31) than condition a (mean=6.32, SD=0.5) and the difference was highly significant in the paired-samples *t*-test ($t1(9)=9.010$, $p<.001$; $t2(19)=11.210$, $p<.001$). Both conditions c and d received very high plausibility ratings (c: mean=6.15, SD=0.8; d: mean=6.36, SD=0.59). Paired-samples *t*-tests revealed that the difference between the two conditions was not significant by item, but were approaching significance by subject ($t1(9)=-2.188$, $p=.056$, $t2(19)=-1.286$, $p>.1$). Similarly, both conditions e and f received high ratings (e: mean=6.03, SD=1.11; f: mean=6.2, SD=0.8), although the difference between the conditions turned out to be marginally significant by subject ($t1(9)=-2.193$, $p=.056$, $t2(19)=-.879$, $p>.1$). The trend of differences in plausibility ratings between conditions c and d and those between e and f needs to be considered when the results of the online task at and after the actual gap positions are interpreted. Importantly, however, the results of the plausibility test show that the plausibility manipulation at the earliest gap site was appropriate, which was crucial for the primary purpose of the present study.

The grammaticality judgment task consisted of 20 items of grammatical and ungrammatical sentences, as shown in (3a)-(3b). The sentences were modified versions of the experimental items of the online task. In the Grammatical condition, the gap, marked with "___" was not inside the island, whereas the Ungrammatical condition involved a movement from within the island. The gap marking was not shown to the participants at the time of testing.

(3) a. Grammatical

I wonder which publisher the author who wrote the book passionately saw ___ while he was traveling.

b. Ungrammatical

I wonder which book the author who wrote ___ passionately saw the publisher while he was traveling.

The participants were randomly assigned to one of two presentation lists. Each list contained only one sentence of an experimental item and 10 sentences per condition. The experimental items were interspersed with 30 distracter items about various grammatical phenomena (e.g., argument structures, *that*-trace effect, *do*-insertion). Only half of the distracters were grammatical.

4.1.3. Procedure

The participants completed the stop-making-sense task before moving on to the grammaticality judgment task. The online task was given before the offline task because the offline task is likely to involve conscious reflection on grammaticality and that might affect their performance in online comprehension.

The stop-making-sense task was administered using the *Linger* software (Doug Rohde, http://tedlab.mit.edu/~dr/Linger/). In each trial, the first screen showed a row of dashes masking each letter of the sentence to read. The participants read the sentence one region at a time by pressing the spacebar. When they pressed the spacebar, the dashes of the first region were replaced by the corresponding words. When they pressed the spacebar again, the first region was masked by the dashes again and the words of the second region appeared in place of the corresponding dashes, and so on through the sentence. While reading each sentence, the participants were asked to press the stop key (the 'Q' key of the keyboard) as soon as they felt it stopped making sense. The trial ended when the participants either read the last region of a sentence or when they pressed the stop key. The task produced two dependent measures for each region—the record of 'stop' responses (0 if the key was not pressed and 1 if it was pressed, reflecting participants' final plausibility judgment at each region) and the reaction time (the time taken to press either the spacebar or the stop key, reflecting time taken to make the plausibility judgment). The participants were instructed to focus on meaning rather than on form. The practice session conducted at the beginning of the task included both natural sentences (e.g., *The woman bought a muffin in the bakery because she was hungry*.) and sentences that were grammatical but implausible (e.g., *The woman bought a muffin in the bird because she was hungry*.). The test items were randomly presented for each participant at the main session.

The grammaticality judgment task was administered using a web-based interface (SurveyGizmo, http://www.surveygizmo.com). The participants read a sentence presented as a whole, and judged whether it was grammatically possible in English on a 4-point scale (1 =

‘impossible in English’, 4 = ‘possible in English’). Practice items were presented before the main items. The items were arranged in a pseudo-randomized order so that no two experimental items would be adjacent to one another.

4.1.4. Predictions

The grammaticality judgment task was administered in order to examine the L2 learners’ grammatical knowledge of the relative clause/subject island constraint. If the learners know that establishing a *wh*-dependency across a relative clause/a subject results in an ungrammatical representation, they will give a significantly lower rating to the Ungrammatical condition than to the Grammatical condition.

The goal of the stop-making-sense task was to show the role of island constraints in real time L2 filler-gap processing, that is, whether islands successfully prevent the active gap search strategy when using the strategy results in an illicit representation. The Non-island conditions were designed to show the time course of *wh*-dependency formation in sentences without an island. Three critical regions (regions 4 to 6) were examined. Region 4 (*wrote passionately* in (1)), which contains the earliest possible gap position and the adverb as a potential spill-over word, was expected to reveal whether the participants actively create a gap. The next region (region 5) contains the preposition (*about* in (1)). Although this is the actual subcategorizer, an initially formed *wh*-dependency might not be abandoned at this point, given that a grammatical continuation without reanalysis is also possible (for at least a subset of the experimental items) as in *I wonder which book the author wrote passionately about his utopia*. It is at region 6 (*while he* in (1)), when it becomes unambiguously clear that there is no overt argument for the preposition, and that the filler should be reinterpreted as the object of the preposition. It was therefore predicted that in case a reanalysis occurs, it might occur at region 6 instead of at region 5. Comparing the Island conditions with the Non-island conditions at these regions was expected to reveal the role of island constraints in filler-gap processing.

Specific predictions are as follows. Numbers 1 and 2 below are in line with the predictions made in Traxler and Pickering (1996). Another possibility is also discussed in number 3.

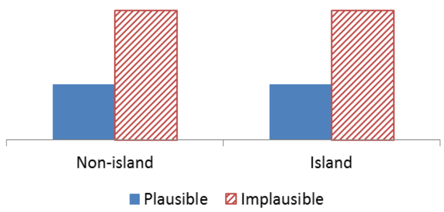
1. If the grammar (in the form of island constraints) does not constrain online filler-gap dependency formation

1) Region 4: Participants will form a wh-dependency upon encountering the first available subcategorizer (*wrote*), due to an urge to form a wh-dependency as soon as possible, regardless of whether the dependency is formed within an island or not. Evaluating an implausible dependency (in the Implausible conditions) will cause higher proportions of stop responses and/or longer reaction times than evaluating a plausible dependency (in the Plausible conditions), and this will happen in both the Non-island and Island conditions. Statistically, this will result in a significant main effect of Plausibility, in the absence of significant interaction with Island.

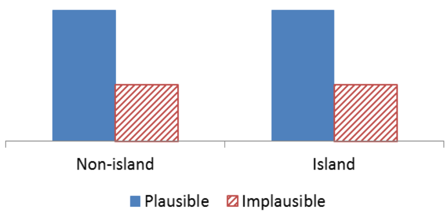
2) Region 5 or 6: The (missing argument of the) next subcategorizer (*about*) provides a cue that reanalysis is necessary. If participants attempt to discard the initially formed wh-dependency and form a new dependency at this position, discarding a plausible initial analysis is likely to be more difficult than discarding an implausible initial analysis. Therefore, there will be more stop responses and/or longer reaction times in the Plausible conditions than in the Implausible conditions, again both in the Non-island and Island

conditions. This will also result in a significant main effect of Plausibility, in the absence of interaction with Island. Figure 4.1 visualizes the predictions.

Figure 4.1. The pattern of stop responses/reaction times predicted if island constraints do not constrain online wh-dependency formation: (a) at region 4 and (b) at region 5 or 6 (if reanalysis is attempted)



(a) *wrote passionately*



(b) *about/saw while he*

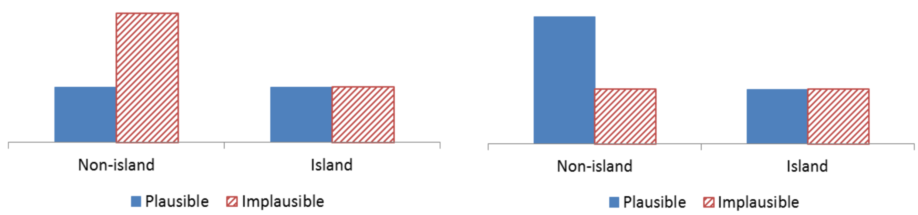
2. By contrast, if the grammar (in the form of island constraints) restricts online filler-gap dependency formation

1) Region 4: Upon encountering the first available subcategorizer, participants will form a wh-dependency, but only in the Non-island conditions. Therefore, in the Non-island conditions, the Implausible condition will be judged as more implausible and/or will be read longer than the Plausible condition, whereas in the Island conditions, plausibility manipulation will not have an effect. Statistical analysis will produce a significant interaction between Plausibility and Island.

2) Region 5 or 6: In case participants attempt to perform a reanalysis upon encountering the (missing argument of the) actual subcategorizer, this will occur only in the Non-island conditions, because in the Island conditions a wh-dependency had not been established at the first available gap site in the first place. This will lead to higher proportions of stop responses and/or longer reaction times in the Plausible than the Implausible conditions

only in the Non-island conditions. Statistically, there will be a significant interaction between Plausibility and Island.

Figure 4.2. The pattern of stop responses/reaction times predicted if island constraints restrict online wh-dependency formation: (a) at region 4 and (b) at region 5 (if reanalysis is attempted)



(a) *wrote passionately* (b) *about/saw while he*

The predictions made in 2 have an underlying assumption that in case the grammar is applied in filler-gap processing, it has a deterministic influence in whether to form a wh-dependency or not. However, it is in principle possible that in addition to the grammatical information (island constraints), plausibility information (plausibility of the wh-dependency to be formed) also affects the parser's decision on whether to posit a gap. Under the assumption that grammar and plausibility work in parallel in online wh-dependency formation, predictions such as 3 are possible instead of 2. For simplicity, it will be assumed below that grammar and plausibility have equal weight in parsing decisions.

3. If both the grammar (in the form of island constraints) and plausibility information affect online filler-gap dependency formation in parallel

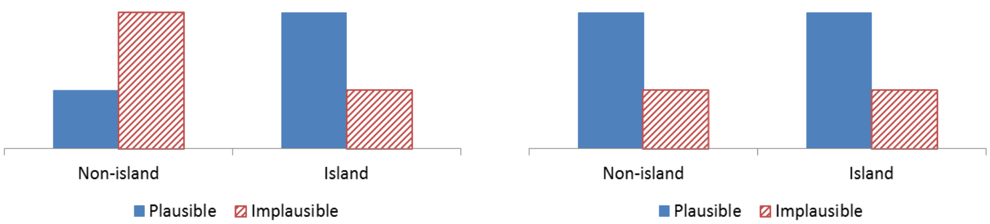
1) Region 4: In the Non-island conditions, island constraints and plausibility support the same analysis in the Plausible condition (both types of information allows the gap

analysis), but they suggest competing analyses in the Implausible condition (island constraints allows the gap analysis but plausibility disprefers it). Therefore, it will be more difficult to determine whether to posit a gap in the Implausible than in the Plausible condition, which will result in a higher proportion of stop response/longer reaction time in the Implausible condition. In the Island conditions, the opposite pattern will occur. It is in the Plausible condition where competition between grammatical information and plausibility information is expected (island constraints prohibits the gap analysis, which is nonetheless plausible), whereas in the Implausible condition, the grammar and plausibility work together to prevent the gap analysis. Determination of whether to posit a gap will be more difficult in the Plausible than in the Implausible condition and the Plausible condition will induce a higher proportion of stop response/longer reaction time. Consequently, the statistical analyses will produce a significant interaction between Plausibility and Island.

2) Region 5 or 6: If participants attempt to reanalyze the initial gap analysis in the Non-island conditions, the Plausible condition will be more difficult to reanalyze (therefore will cause a higher proportion of stop response/a longer reaction time) than the Implausible condition. This is because the initial gap analysis had been supported by more information in the Plausible condition (by both the grammar and plausibility) than in the Implausible condition (only by the grammar), and the participants are more likely to have committed to the initial gap analysis more strongly when more information had supported it. In the Island conditions, it is also the Plausible condition that is more likely to induce higher proportions of stop responses/longer reaction times than the Implausible condition. The initial gap analysis had been supported by one type of information in the

Plausible condition (by plausibility). Therefore, under the assumption that the grammar and plausibility have an equal influence, the parser is predicted to have considered the initial gap analysis to a certain degree in this condition, which might produce some degree of difficulty in reanalysis. In the Implausible condition, in which neither the grammar nor plausibility had supported the initial gap analysis, there will be no reanalysis. This will result in a main effect of Plausibility in the statistical analysis, without an interaction with Island.

Figure 4.3. The pattern of stop responses/reaction times predicted if both island constraints and plausibility affect online wh-dependency formation: (a) at region 4 and (b) at region 5 or 6 (if reanalysis is attempted)



(a) *wrote passionately* (b) *about/saw while he*

Note that the predicted pattern at region 5/6 is similar in Prediction 1 and 3, although the mechanisms causing the pattern are not identical. Therefore, the overall results of all critical regions need to be taken into account to infer the processes involved.

4.2. Results

4.2.1. Grammaticality Judgment Task

The grammaticality judgment task produced ratings with four ordinal numbers from 1 to 4, with higher number indicating higher grammaticality. The mean ratings given to the Grammatical and Ungrammatical conditions by each group were submitted to paired-samples t-tests to examine whether they were reliably different. The ratings given by the two groups in each condition are summarized in Table 4.2.

Table 4.2. Means and standard deviations of the grammaticality ratings (SD in parenthesis)

	Grammatical	Ungrammatical
Native speakers	3.07 (1.16)	1.28 (0.69)
L2 learners	3.58 (0.91)	2.00 (1.30)

The grammaticality ratings show that both the native speakers and the L2 learners clearly distinguished the sentences that violated island constraints from their grammatical counterparts. Paired-samples t-tests revealed that both groups gave a significantly lower rating to the Ungrammatical condition than to the Grammatical condition (native speakers: $t1(23)=10.624$, $p<.001$; $t2(19)=11.721$, $p<.001$; L2 learners: $t1(30)=8.342$, $p<.001$; $t2(19)=15.215$, $p<.001$). The learners in this study therefore appear to have acquired the grammar of the relative-clause/subject island constraint, which can potentially be applied in online comprehension.

4.2.2. Stop-Making-Sense Task

Plausibility judgments

Table 4.3 presents a summary of stop responses at each critical region. Proportions of stop response were calculated relative to the number of possible responses in each condition at each region, following Williams et al. (2001). For example, if a participant pressed the stop key in 1 out of 5 sentences of a condition at region N, the proportion rate was 0.2. At region N+1, the number of sentences belonging to the same condition was 4, because the sentence in which the participant had pressed the stop key at the previous region is no longer presented. If the participant pressed the stop key in one sentence at region N+1, the proportion of stop response of this region was thus 0.25.

Table 4.3. Mean proportions of stop responses and standard deviations at the critical regions: Native speakers and L2 learners

	Region 4	Region 5	Region 6
Native speakers			
Non-island/plausible	0.01 (0.09)	0.01 (0.09)	0.19 (0.39)
Non-island/implausible	0.31 (0.46)	0.04 (0.19)	0.08 (0.27)
Island/plausible	0.08 (0.28)	0.03 (0.16)	0.05 (0.21)
Island/implausible	0.09 (0.28)	0.04 (0.19)	0.07 (0.25)
L2 learners			
Non-island/plausible	0.04 (0.19)	0.02 (0.14)	0.05 (0.21)
Non-island/implausible	0.19 (0.40)	0.02 (0.13)	0.07 (0.25)
Island/plausible	0.04 (0.20)	0.06 (0.24)	0.05 (0.22)
Island/implausible	0.03 (0.18)	0.04 (0.20)	0.01 (0.12)

To examine whether the observed effects are statistically reliable, the data from each group were submitted to logistic mixed effects modeling in each critical region. For all models, Plausibility and Island, and their interactions were included as fixed effect terms, and participants and items as random effect terms. Plausibility (Plausible vs. Implausible) and Island (Non-island vs. Island) were coded using mean-centered contrast coding. In case a significant interaction between Plausibility and Island is found, the source of the interaction was explored by fitting additional mixed effects models in the Non-island conditions and in the Island conditions, separately. Table 4.4 presents the fixed effects of the model fitted to the data from native speakers in each region. The estimates represent the magnitude of the corresponding fixed effects in log odds. For example, the estimate for Plausibility at region 4, 2.301, indicates that the odds of stop response was 9.98 (i.e., exp(2.301)) times higher in the implausible than in the plausible conditions.

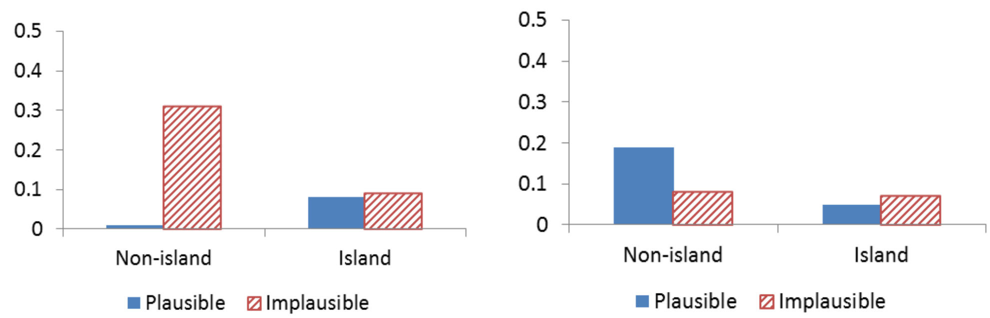
Table 4.4. Fixed effects of the model fitted to the judgment data in each region: native speakers

		estimate	Std. error	z value	p value
Region 4	(Intercept)	-3.211	0.434	-7.392	0.001 ***
	Plausibility	2.301	0.628	3.665	0.001 ***
	Island	0.292	0.631	0.462	0.644
	Plausibility:Island	-4.540	1.254	-3.619	0.001 ***
Region 5	(Intercept)	-4.059	0.437	-9.282	0.001 ***
	Plausibility	0.897	0.765	1.172	0.241
	Island	0.654	0.800	0.818	0.414
	Plausibility:Island	-1.172	1.557	-0.753	0.452
Region 6	(Intercept)	-2.952	0.368	-8.033	0.001 ***
	Plausibility	-0.295	0.430	-0.684	0.494
	Island	-1.026	0.426	-2.408	0.016 *
	Plausibility:Island	1.645	0.870	1.890	0.059 †

* The number of asterisks besides the p values indicates the level of significance at 0.001***, 0.01**, 0.05*, 0.1†

At region 4, which contained the earliest possible gap site, the analysis revealed a significant main effect of Plausibility, due to a higher rate of stop response in the Implausible than the Plausible conditions, but the effect of Plausibility was modulated by the Island factor. In order to further examine the source of interaction between Plausibility and Island, mixed-effects logit models were fitted to the data from the Non-island and Island conditions, separately. The results showed that the plausibility effect was significant in the Non-island conditions, in the direction of more stop responses in the Implausible than in the Plausible condition (estimate=4.098, SE=1.079, z=3.797, p<.001), but the plausibility effect was not significant in the Island conditions (estimate=0.070, SE=0.587, z=0.120, p>.1). At region 5, there were no significant main effects or interaction. At region 6, the main effect of Island was significant, due to a higher number of stop responses in the Non-island compared to the Island conditions. The interaction between Plausibility and Island almost reached significance. Mixed effects logit models separately fitted to the data from the two Island conditions revealed that the plausibility effect was marginally significant in the Non-island conditions, due to a higher number of stop responses in the Plausible than in the Implausible condition (estimate=-1.008, SE=0.527, z=-1.912, p=.056), whereas there was no significant plausibility effect in the Island conditions (estimate=2.833, SE=3.086, z=0.918, p>.1). As visualized in Figure 4.4 below, the results overall fits prediction 2 above.

Figure 4.4. Proportions of stop response of native speakers (a) at region 4 and (b) at region 6



(a) *wrote passionately* (b) *while he*

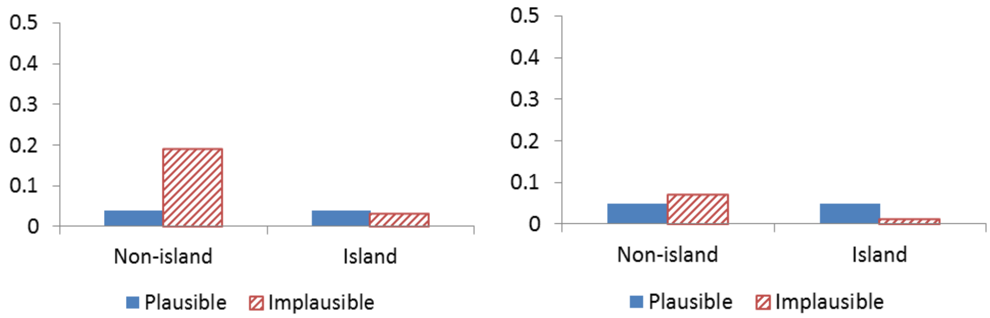
Turning to the L2 learners' data, the fixed effects of the model fitted to the data from each region is presented in Table 4.5.

Table 4.5. Fixed effects of the model fitted to the judgment data in each region: L2 learners

		estimate	Std. error	z value	p value
Region 4	(Intercept)	-3.548	0.357	-9.935	0.001 ***
	Plausibility	0.919	0.421	2.181	0.029 *
	Island	-1.078	0.422	-2.554	0.011 *
	Plausibility:Island	-2.311	0.842	-2.743	0.006 **
Region 5	(Intercept)	-3.615	0.299	-12.080	0.001 ***
	Plausibility	-0.319	0.555	-0.575	0.565
	Island	1.047	0.560	1.869	0.061 †
	Plausibility:Island	-0.301	1.134	-0.265	0.791
Region 6	(Intercept)	-3.573	0.322	-11.111	0.001 ***
	Plausibility	-0.527	0.534	-0.987	0.324
	Island	-0.718	0.517	-1.390	0.165
	Plausibility:Island	-1.772	1.058	-1.674	0.094 †

At region 4, there were significant main effects of Plausibility and Island, due to a higher rate of pressing the stop key in the Implausible than in the Plausible conditions, and in the Non-island than in the Island conditions, respectively. The interaction between Plausibility and Island was significant. A subsequently conducted mixed effects logit model fitted to the two island conditions separately showed that the plausibility effect was significant in the Non-island conditions (estimate=2.258, SE=0.545, z=4.146, $p<.001$) whereas it was not in the Island conditions (estimate=-0.210, SE=0.617, z=-0.340, $p>.1$). At region 5, only the main effect of Island was marginally significant. At region 6, there was a trend toward an interaction between Plausibility and Island. This was because the rate of stop responses was numerically higher in the Implausible conditions within the Non-island conditions, but in the Plausible conditions within the Island conditions, a pattern not predicted at this region. Subsequently performed mixed effects logit models, however, revealed no significant plausibility effect either in the Non-island conditions (estimate=0.712, SE=0.690, z=1.032, p>.1) or in the Island conditions (estimate=-1.349, SE=0.855, z=-1.577, p>.1).

Figure 4.5. Proportions of stop response of L2 learners (a) at region 4 and (b) at region 6



(a) *wrote passionately* (b) *while he*

To sum up, the results of the plausibility judgment show that native speakers attempted a filler-gap dependency at the earliest possible gap position, and were sensitive to the information signaling that this initially formed filler-gap dependency should be reanalyzed. Moreover, active gap search and the subsequent reanalysis occurred only when the earliest possible gap position was not within an island, suggesting that native speakers apply island constraints in online filler-gap processing. Like native speakers, L2 learners formed a wh-dependency at the earliest possible gap site, although when this turned out to not to be the correct parse, they seemed to have more difficulty with reanalysis than native speakers. Crucially, however, the learners did not use the active gap search strategy when the gap site was within an island. L2 learners, therefore, showed evidence of obeying grammar when forming a filler-gap dependency online.

Reaction times

The reaction time of each segment was recorded while the participants conducted the online task. In case participants pressed the stop key while reading a sentence, reaction times were available up to the region on which they pressed the stop key because the sentence was not presented after that point. Reaction times that were above or below 2.5 standard deviations away from the group mean at each region in each condition were removed, and this affected 3.61% of the data. Table 4.6 presents the mean reaction time and standard deviation calculated from the remaining data in each condition for each language group.

Table 4.6. Means and standard deviations of the reaction times of the critical regions: Native speakers and L2 learners

	Region 4	Region 5	Region 6
Native speakers			
Non-island/plausible	1001 (444)	618 (245)	944 (601)
Non-island/implausible	1160 (548)	642 (220)	652 (283)
Island/plausible	1221 (709)	681 (282)	648 (283)
Island/implausible	1136 (514)	738 (341)	668 (287)
L2 learners			
Non-island/plausible	1852 (1173)	695 (306)	1101 (697)
Non-island/implausible	2028 (1060)	714 (327)	1052 (583)
Island/plausible	1941 (988)	1180 (839)	1029 (636)
Island/implausible	2081 (1086)	1126 (726)	1114 (775)

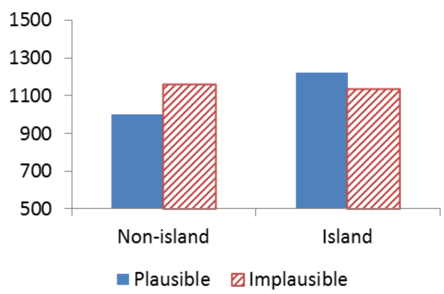
The data of each group were submitted to linear mixed effects modeling at each critical region, with Plausibility and Island as the fixed effect terms and participants and items as the random effect terms. P-values for the fixed effects were derived through parameter estimation based on Markov chain Monte Carlo (mcmc) sampling. Table 4.7 shows the fixed effects of the models fitted to the data from native speakers in each region.

Table 4.7. fixed effects of the model fitted to the reaction times in each region: native speakers

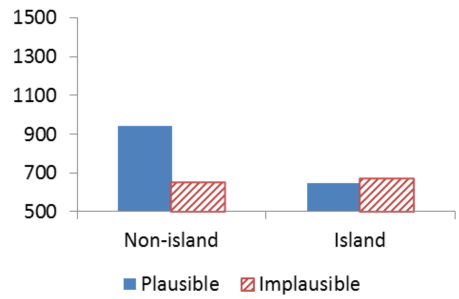
		estimate	Std. error	t value	p value
Region 4	(Intercept)	1132.15	64.50	17.553	0.001 ***
	Plausibility	36.24	45.80	0.791	0.429
	Island	93.50	45.96	2.034	0.043 *
	Plausibility:Island	-263.72	91.98	-2.867	0.004 **
Region 5	(Intercept)	673.32	30.36	22.176	0.001 ***
	Plausibility	37.52	24.73	1.517	0.136
	Island	72.96	24.60	2.966	0.003 **
	Plausibility:Island	18.56	49.48	0.375	0.679
Region 6	(Intercept)	744.31	41.73	17.837	0.001 ***
	Plausibility	-130.82	38.19	-3.425	0.001 ***
	Island	-156.13	38.31	-4.075	0.001 ***
	Plausibility:Island	297.17	77.19	3.850	0.001 ***

At region 4, the main effect of island was significant, due to longer reaction times in the Island conditions than in the Non-island conditions, which is not surprising considering that the Island conditions had one more word than the Non-island conditions (*who wrote passionately* vs. *wrote passionately*). There was also significant interaction between Plausibility and Island. Linear mixed effects models fitted to the data from the Non-island and Island conditions separately revealed that the reaction time was significantly longer in the Implausible than in the Plausible conditions in the Non-island condition (estimate=160.07, SE=60.25, t=2.657, p<.01), but there was no reaction time difference between the two plausibility conditions in the Island conditions (estimate=-92.59, SE=68.37, t=-1.354, p>.1). At region 5, only the main effect of Island was significant, due to longer reaction times in the Island than in the Non-island conditions. At region 6, there were significant main effects of Plausibility and of Island, due to longer reaction times in the Plausible conditions and in the Non-island conditions, respectively. The interaction between Plausibility and Island was also significant. Linear mixed effects models subsequently fitted to the two island conditions showed that the plausibility effect was significant in the Non-island conditions, in the direction that Plausible conditions had longer reaction times than the Implausible conditions (estimate=-287.66, SE=67.79, t=-4.243, p<.001), but the plausibility effect was not significant in the Island conditions (estimate=11.15, SE=35.12, t=0.318, p>.1).

Figure 4.6. Reaction times of native speakers (a) at region 4 and (b) at region 6



(a) *wrote passionately*



(b) *while he*

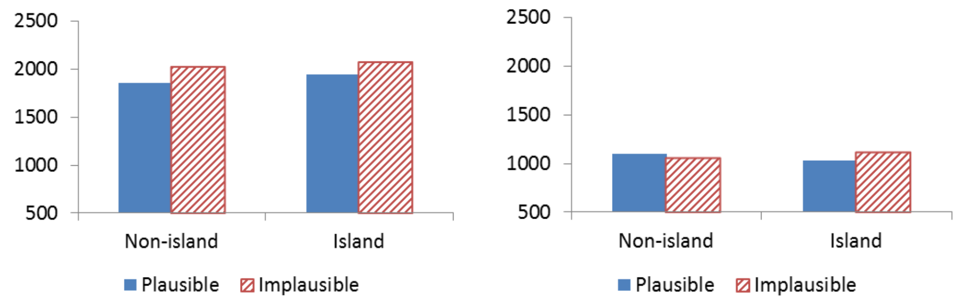
Table 4.8 shows the fixed effects of the model fitted to the learners' data.

Table 4.8. fixed effects of the model fitted to the reaction times in each region: L2 learners

		estimate	Std. error	t value	p value
Region 4	(Intercept)	1978.683	95.288	20.765	0.001 ***
	Plausibility	167.972	81.302	2.066	0.039 *
	Island	86.101	81.506	1.056	0.291
	Plausibility:Island	-2.965	162.911	-0.018	0.986
Region 5	(Intercept)	947.132	50.885	18.613	0.001 ***
	Plausibility	-3.725	48.547	-0.077	0.939
	Island	453.377	48.648	9.320	0.001 ***
	Plausibility:Island	-83.303	97.427	-0.855	0.393
Region 6	(Intercept)	1068.430	49.200	21.716	0.001 ***
	Plausibility	10.545	57.031	0.185	0.853
	Island	4.665	0.082	0.082	0.935
	Plausibility:Island	147.244	114.382	1.287	0.199

At region 4, there was a significant main effect of Plausibility, due to longer reaction times in the Implausible conditions than in the Plausible conditions. Importantly, Plausibility did not interact with Island, unlike in native speakers. At region 5, there was a significant main effect of Island, due to longer reaction times in the Island than in the Non-island conditions. At region 6, there were no significant main effects or interaction.

Figure 4.7. Reaction times of L2 learners (a) at region 4 and (b) at region 6



(a) *wrote passionately* (b) *while he*

To summarize the results of reaction times, the pattern of native speakers' reaction times was consistent with the pattern of their plausibility judgments. There was an active gap search effect and a reanalysis effect, but only when the earliest possible gap site was not within an island. The pattern of L2 learners' reaction times, however, was quite different from the pattern of their plausibility judgments, in that there was an active gap search effect, but regardless of whether the earliest possible gap site is within an island or not. The reanalysis effect did not occur in the reaction times as well.

4.3. Discussion

In Experiment 1, an offline grammaticality judgment task and an online stop-making sense task were administered to adult Korean-speaking learners of English with advanced proficiency. The study investigated the learners' grammatical knowledge of the relative clause/subject island constraint and the way they make use of the constraint in processing wh-dependencies.

The grammaticality judgment task showed that the L2 learners gave a significantly lower grammaticality rating to ungrammatical sentences, which suggests that they have grammatical knowledge of the island constraint. The literature on L2 acquisition of island constraints have produced a large body of evidence that adult L2 learners with a wh-*in situ* L1 background are not sensitive to island constraints in offline grammaticality judgment tasks (e.g., Hawkins & Hattori, 2006; Johnson & Newport, 1991; Schachter, 1990; cf, White & Juffs, 1998). Contrary to these studies, the results of Experiment 1 clearly show that adult L2 learners with a wh-*in situ* L1 can assign proper grammaticality ratings to the sentences that violated the relative clause/subject island constraint and to those that did not violate the constraint. This finding provides a challenge to the view that there is maturational decline in the ability to access island constraints as a subset of Universal Grammar in L2 acquisition (Johnson & Newport, 1991).

In the stop-making-sense task, the baseline sentences were those that did not include an island (the 'Non-island' conditions). These sentences made it possible to examine two phenomena in filler-gap processing that occur when a potential wh-dependency to be formed does not cross an island boundary: 1) whether the parser actively searches for the gap and 2) whether the parser reanalyzes a previously formed (but incorrect) wh-dependency. As well-established in the L1 processing literature (e.g., Crain & Fodor, 1985; Clifton & Frazier, 1989; Garnsey et al., 1989; Stowe, 1986; Traxler & Pickering, 1996), both implausibility detection rates and reaction times showed that native speakers formed a wh-dependency at the earliest possible gap site. The tendency to actively search for a gap was strong enough to override the effect of the plausibility of the filler-gap association. Native speakers formed a wh-dependency with the first verb even when the filler was not a plausible object of this verb. Like native speakers, the L2 learners also attempted to search for the gap of the clause-initial filler at the earliest possible gap site to the extent that the effect of plausibility is overridden. Experiment 1 thus adds another piece of evidence to the growing body of studies showing that L2 filler-gap processing strongly relies on active gap search strategy (Aldwayan et al., 2010; Dussias & Pinar, 2010; Hoover & Dwivedi, 1998; Williams et al., 2001). The parser does not wait to encounter a necessary-but-missing argument of a subcategorizer before positing a gap. It instead appears to actively predict and anticipate where the gap might be and incrementally processes filler-gap dependencies in L2. The results of the 'Non-island' sentences therefore show that L1 and L2 processing use a qualitatively similar—perhaps a universal—processing strategy.

In the Non-island sentences, the wh-dependency formed at the earliest possible gap site turned out not to be the correct dependency. The parser therefore faced the task of revision, which involves nullifying the already formed wh-dependency with the first verb and completing a new dependency between the filler and the preposition. The experiment examined whether or not the parser revised the initial parse, by using the 'reanalysis effect' as a barometer (the relative difficulty in processing the sentences with an initially plausible dependency than processing those with an initially implausible dependency). The native speakers produced a reanalysis effect in both implausibility detection rates and reaction times at region 6, suggesting that upon encountering the missing argument of the preposition, they realized that the dependency with the first verb was not the correct analysis and reanalysis was necessary. It needs to be pointed out, however, that although the results indicate that the native speakers were sensitive to the cue for reanalysis and attempted at revision, it is not clear whether they completely succeeded in it. In fact, a completely successful revision would cause a reanalysis effect in reaction times (because it would take more time to reanalyze a plausible initial parse than to reanalyze an implausible initial parse) but not in plausibility judgment (because all sentences were ultimately plausible at and after the ultimate gap site). It appears that reanalyzing a plausible initial parse was so costly that native speakers sometimes mistakenly judged the sentences as ultimately implausible. This is in line with previous L1 studies showing that native speakers are not always successful in reanalysis (e.g., Christianson et al., 2001; Ferreira & Henderson, 1991).

In the offline plausibility test conducted before the main session (example (2)), condition c received a lower rating than d (means: c = 6.15, d = 6.36 out of the maximum of 7) and similarly e received a lower rating than f (means: e = 6.03, f = 6.2). Although the numerical differences were small, they were marginally significant in the subject analysis, and the possibility cannot be excluded that the 'reanalysis effect' produced in the online task might actually reflect that the sentences of the Plausible conditions were slightly less plausible at the actual gap position than the sentences in the Implausible conditions. This does not seem very likely. In the offline plausibility test, the Plausible condition received a lower rating than the Implausible condition in both the Non-island conditions and the Island conditions. If differences in plausibility had an effect on the online task, a 'reanalysis effect' must have occurred in both the Non-island and the Island conditions, but the 'reanalysis effect' occurred in the Non-island conditions only. Therefore, it seems reasonable to interpret that the 'reanalysis effect' is indeed an indicator of (attempted) revision.

Unlike the native speakers, the learners did not show a reanalysis effect either in implausibility detection rates or in reaction times. Since there was no reanalysis effect at all, the simplest explanation of this result seems to that the L2 learners had difficulty with reanalysis and the difficulty was more severe compared to the native speakers. Although the lack of a reanalysis effect in implausibility detection rates could in principle reflect a successful completion of reanalysis as discussed above, the L2 learners' results do not appear to suggest that they correctly reparsed the sentences, given that they did not produce a reanalysis effect in reaction times. Overall, the results of the Non-island conditions suggest that both L1 and L2 filler-gap processing rely on a default strategy of forming the shortest-distance wh-dependency but the two may differ in the extent to which they can recover if this initially formed dependency turns out to be incorrect.

The role of island constraints in filler-gap processing was examined by comparing the Island conditions to the baseline Non-island conditions. If island constraints do not have an immediate effect in online filler-gap dependency formation, the parser will create a wh-dependency with the first encountered subcategorizer, regardless of whether the dependency spans an island boundary or not. If, on the other hand, island constraints immediately set a limit on parsing decisions so that only grammatical wh-dependency is formed, the parser will suspend the usual active gap search in an island domain. Consistent with the findings of many previous L1 processing studies (e.g., McElree & Griffith, 1998; Stowe, 1986; Traxler & Pickering, 1996), the native speakers of Experiment 1 showed evidence of suppressing active gap search while reading the island domain, in both plausibility judgments and reaction times. The routine of completing a wh-dependency as soon as possible appears to have been inhibited for the sake of obeying grammatical constraints.

The pattern of the L2 learners' plausibility judgment was similar to that of the native speakers. There was an 'active gap search effect' in the Non-island conditions only, a pattern indicating that at the earliest possible gap site, the learners decided to posit a gap in the Non-island conditions but not in the Island conditions. It suggests that the learners had successfully applied island constraints by the time they made the final parsing decision at the earliest possible gap site. A qualitatively different pattern was found in the reaction time, however, which is a measure of the time taken to judge whether a given sentence is plausible or not at a certain region. There was evidence of active gap creation, but whether or not the earliest possible gap position was within an island did not have an effect. Assuming that the reaction time of a region reflects the processes the parser goes through at the region until it makes the final parsing decision, the results could be taken to indicate that the learners have entertained the ‘earliest gap’ analysis, regardless of whether it is grammatically licit or not, at some point in processing before finally settling on a grammatical parse. Comprehending L2 sentences imposes more processing burden than comprehending sentences in one’s native language. Therefore, it may be difficult for the L2 learners to completely suppress the urge to rapidly complete an unfinished wh-dependency—which is probably a memory-friendly parsing option—within an island. Although the results of Experiment 1 suggest that L2 learners may be slower and less efficient in applying island constraints, what is clearly shown is that the learners are capable of applying the constraints in real time processing to ultimately rule out ungrammatical representation and compute structural representations that obeys grammatical constraints in L2.

Chapter 5

Experiment 2

Experiment 1 investigated the way Korean learners of English use the subject/relative clause island constraint in online filler-gap processing using the stop-making-sense task. In Experiment 1, native English speakers showed evidence of immediate island sensitivity: they suspended the default strategy of forming the shortest-possible wh-dependency while processing a relative clause modifying a subject NP, which is an extraction island. Korean learners of English were less efficient than native speakers in using the subject/relative clause island constraint and were not able to fully inhibit the inclination to attempt the shortest-possible wh-dependency within the island. However, they were able to ultimately apply the island constraint and reject the ungrammatical wh-dependency.

The primary goal of Experiment 2 was twofold. The first purpose was to examine sensitivity to island constraints using another online technique. Although the stop-making-sense can provide insights on language processing, it is not without some limitations as a tool to examine normal language processing. First of all, it forces participants to explicitly judge the plausibility of what they are reading. Since the participants keep monitoring the sentences, the psycholinguistic processes involved might not be identical to the processes occurring when the only purpose of reading is comprehension. Moreover, words were presented one region at a time in the non-cumulative fashion in Experiment 1, as most studies using the stop-making-sense (or self-paced reading) task did. This way of sentence presentation could possibly encourage readers to process sentences more incrementally than when entire sentences are presented. It therefore remains to be seen whether the participants would be sensitive to island constraints while performing an online task that more closely approximates typical reading situations. To address this issue, the participants of Experiment 2 were allowed to read sentences, which were presented as a whole, at their own pace and in their own way, without being asked to explicitly monitor any aspect of the sentences. While they were reading, their eye movements were monitored using an eye-tracker. Experiment 2 therefore investigated whether participants apply island constraints in an ecologically more valid (albeit not entirely identical) reading situation. Converging evidence from different methodologies would make it possible to make a more robust claim that the learners are able to use island constraints in online processing.

The second purpose of the present experiment was to investigate the effect of individual differences in cognitive resources on readers' sensitivity to island constraints in L1 and L2 sentence processing. Although individual differences in the sensitivity to island constraints have not yet been paid much attention in previous work on L1 and L2 sentence processing, one factor that was associated with the operation of island constraints is individual differences in working memory capacity (e.g., Hofmeister & Sag, 2010; Kluender & Kutas, 1993). The potential role of working memory capacity in processing island constraints has been entertained under the view that the island phonemena are not due to grammaticalized constraints but to the processing difficulty incurred by island constructions. According to this view, a reader's working memory capacity might determine his or her sensitivity to island constraints; specifically, it is predicted that readers with 'smaller' working memory capacity will produce a 'stronger' island effect. Under this view, island effects occur because the processing complexity of island constructions drains processing resources and leaves no room for the filler to be reactivated. For the readers with larger working memory capacity, it is easier to manage to form a filler-gap dependency even within island constructions. For those with smaller working memory capacity, however, forming a filler-gap dependency within an island is difficult. This prediction has not been borne out in a recent study on English native speakers' offline acceptability judgments of island constraints that suggests that individual differences in working memory capacity do not reliably constrain the appreciation of island constraints at least in the phase of acceptability judgments following online sentence processing (Sprouse et al., 2012). It has not yet been tested, however, whether individual differences in working memory capacity are associated with individual differences in the sensitivity to island constraints during online sentence processing. A better understanding of the effect of working memory capacity on online island sensitivity will be able to offer insights regarding the question of whether the island phenomena are the result of mentally represented grammatical constraints or that of the capacity-limited nature of human processing resources. It is also important for L2 research because whether or not the island constraint is a syntactic phenomenon has consequences for the debate on the depth of syntactic processing in L2 (Cunnings et al., 2010; Omaki & Schulz, 2011). Moreover, the exploration of the role of working memory in L2 learners' sensitivity to island constraints will help us better understand the role of working memory capacity in general L2 sentence comprehension processes (Dussias & Pinar, 2010; Felser & Roberts, 2007; Havik et al, 2009; Juffs, 2004; McDonald, 2006).

To address the research goals described above, Experiment 2 monitored eye-movement patterns of Korean learners of L2 English while they process filler-gap sentences with and without islands, and also investigated the effect of individual differences in working memory capacity on their sensitivity to island constraints during online sentence processing.

5.1. Method

5.1.1. Participants

Language background and English proficiency

Nineteen native speakers of English (age 19-23, mean = 20.6) and 23 Korean learners of English (age 22-33, mean = 27.1) who did not participate in Experiment 1 participated in Experiment 2. All participants were recruited from the University of Illinois at Urbana-Champaign community. The L2 learners completed the same language background questionnaire and the cloze test used for Experiment 1, and the results are summarized in Table 5.1. For ease of comparison, the language backgrounds and the close test scores of the learners of Experiment 1 are presented together. Table 5.1 shows that the learners of the two experiments were generally similar in terms of their English proficiency and language backgrounds. Independent samples *t*-tests were conducted to compare the learners of the two experiments on each measure. The two groups were not different in the cloze test scores ($t(50)=1.230$, $p>.1$), suggesting that they had comparable general English proficiency. The two groups were also not different in the age at which they were first taught English in Korea ($t(50)=-.224$, $p>.1$), in the age at which they arrived in the US ($t(50)=1.169$, $p>.1$) and in the percentage of using English per day ($t(50)=.399$, $p>.1$). The only significant difference were in the total duration of stay in the US at the time of testing, with the learners of Experiment 2 having lived in the US longer than those of Experiment 1 ($t(50)=-2.089$, $p<.05$).

Table 5.1. L2 learners' cloze test scores and language background information in Experiment 1 and 2

		cloze test	age of first	age of first	duration of	English use
Exp 1 (N=31)	range	28-37	7-14	15-34	Less than 1-8	5-100
	mean	33.6	11.1	23.1	3.6	40.8
	SD	2.4	1.9	5.1	2.4	26.8
Exp 2 (N=22)	range	27-38	6-15	15-31	Less than 1-9	5-70
	mean	32.8	11.2	21.6	5.1	38.1
	SD	3.1	2.1	5.7	2.8	18.2

5.1.2. Working memory span

In order to measure participants' individual working memory capacity, two working memory span tests were administered in English. In addition to the reading span test, which is one of the most widely used span tests, the subtract-2-span test was also administered. A composite working memory score was calculated for each participant from the results of the two tests, following Waters & Caplan (2003), who suggest that using a composite measure derived from multiple span tests improves reliability.

Reading span test

In the reading span test, participants were engaged in a dual task of remembering letters while comprehending sentences. The reading span test was a version modified from the original Daneman and Carpenter's (1980) test (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005). The task was administered on a PC. (1a) to (1b) show a series of screens that the participants saw in an example trial.

(1) a. (1st screen) A person should never be discriminated against based on his race. ? M

b. (2nd screen) My mother has always told me that it is not polite to shine. ? L

c. (3rd screen) The lemonade players decided to play two out of three sets. ? F

d. (4th screen)   ???

The screens, except for the last one, showed an English sentence, followed by a question mark and a random letter. The sentences were plausible about half of the time and implausible in the other half. The participants read out the given sentence, and at the question mark indicated whether the sentence makes sense by saying ‘yes’ or ‘no’, and read out the letter next to the question mark (for example, if the screen shows (1a), participants said, “A person should never be discriminated against based on his race. Yes. M”). They then pressed the space bar to see the next screen and repeated the procedure until they see the recall prompt on the next screen, in the form of three question marks (1d). At that point, the participants wrote the letters on the paper provided, in the order that they saw them (in (1) for example, the correct answer would be ‘M, L, F’). After writing the letters, the participants pressed the space bar to move to the next trial. The number of sentence-letter pairs in a trial varied from two to five ((1) is an example of the three sentence-letter pairs). There were three trials for each number of sentence-letter pairs, resulting in twelve trials and forty-two letters to recall in total. The order of trials was randomized such that the participants would not be able to guess how many sentence-letter pairs they would see in a given trial. Participants completed three practice trials, all with two sentence-letter pairs, before proceeding to the main session.

While the participants were conducting the reading span task, the accuracy of the participants’ response on the plausibility of each sentence was recorded by the experimenter. Participants were notified of this fact in advance, to be encouraged to pay attention to the plausibility judgment component. The native speakers’ plausibility judgment scores ranged from

38 to 42 out of the maximum of 42 (all over 90%). The L2 learners' plausibility judgment scores were generally lower than those of the native speakers, with some learners having quite low scores (range=24-40 (57.1%-95.2%, M=33.3 (79.4%), SD=4.8 (11.5%)). However, no L2 learner was excluded from the analysis, considering the possibility that low plausibility judgment scores might have resulted from poor reading comprehension ability, not from lack of attention to sentence comprehension component of the reading span task.

The reading span score of each participant was determined by counting the total number of correct letters in the correct position. The lowest possible score was thus zero and the highest possible score was forty-two.

Subtract-2 span test

For the subtract-2 span test, a sequence of digits was automatically presented on a computer screen one by one and participants read them out as they saw each one. After all digits were presented, blank boxes of the same number of presented digits appeared on the next screen. At that point, participants wrote the digits in the boxes in the order of presentation, after subtracting 2 from each digit. For example, if '5', '2', '9', '3' popped up on the screen one by one, participants said aloud "five", "two", "nine", "three" after they saw each digit. The next screen showed four blank boxes, and the correct response was to write '3', '0', '7' and '1' in each box. The length of a trial varied from two to seven digits and there were two trials for each number of digits, resulting in 12 trials in total and 54 number of digits in total. It was not possible to predict the number of digits in a given trial because the order of trials was randomized for each participant. Participants were given two practice trials at the beginning of the test. Both practice trials included two digits. The number of correctly entered answers was added up to compute the subtract-2 span test score of each participant (lowest possible score=0, highest possible score=54)

Composite working memory score

The reading span test score and the subtract-2-span test score of each participant were converted to percentage respectively. The average of these converted scores of each participant was used as the composite working memory score of the participant. Table 5.2 presents a summary of each group's composite working memory score. The reading span scores and the subtract-2 span scores in percentages are also presented, although these were not directly used in the analysis.

Table 5.2. Native speakers' and L2 learners' working memory scores (%)

		reading span	subtract-2	composite
Natives (N=19)	range	47.6 – 92.9	59.3 – 96.3	59.4 – 88.8
	mean	74.7	80.5	77.6
	SD	12.4	10.0	8.3
L2 learners (N=22)	range	35.7 – 97.6	42.6 – 90.7	50.3 – 89.9
	mean	72.7	72.9	72.8
	SD	16.5	11.9	10.9

A noticeable feature of the results in Table 5.2 is that although the L2 learners received numerically lower average scores than the native speakers in all three scores, the differences were not large. Independent samples *t*-tests conducted to compare the two groups' scores revealed that the native speakers' scores were significantly higher than the L2 learners' scores only in the subtract-2 span test ($t(39)=2.192$, $p<.05$). The two groups were not significantly different in the reading span test score or in the composite working memory score (*p*s>.1).

5.1.3. Materials

For the eye-tracking task, two factors (Island and Plausibility) were crossed to form four conditions as in Experiment 1. Although there were some minor changes in lexical items, the design and sentence structures of the experimental sentences of the eye-tracking task were identical to those of the online task of Experiment 1. An example set of experimental sentences is presented in (2) below and the complete set can be found in Appendix A. Twenty sets of experimental items were distributed to four presentation lists in a Latin Square design. Sixty-four filler sentences were also added to each list. The filler sentences were different from those of Experiment 1. This was because it is desirable to use globally plausible sentences for an eye-tracking task whereas it is desirable to use at least some implausible sentences for a stop-making-sense task. Participants were randomly distributed among the four presentation lists and sentences were presented in a randomized order for each participant.

(2) a. Non-island/Plausible

The reporter asked which book the author wrote passionately about while he was traveling.

b. Non-island/Implausible

The reporter asked which city the author wrote passionately about while he was traveling.

c. Island/Plausible

The reporter asked which book the author who wrote passionately saw while he was traveling.

d. Island/Implausible

The reporter asked which city the author who wrote passionately saw while he was traveling.

A true/false statement was created for about half of the items. For example, for the sentence "*The police learned which hotel the maid cleaned occasionally for before she disappeared.*", the question was "*The police were looking for a missing maid*.". No question directly asked what was investigated in the present experiment (i.e., which subcategorizer the filler should be associated with).

5.1.4. Procedure

Participants sat in front of a computer equipped for EyeLink 1000. After the participants rested their chin on the chinrest and leaned their forehead against the head rest, the eye-tracker was calibrated on a 9-point grid. Viewing was binocular but calibration was performed with only the right eye since only the movement of the right eye was recorded. Four practice trials were given to the participants before the main session in order for them to become familiar with the experimental procedure. Each trial began with a dot at the left corner of the screen in the middle from top to bottom. Participants were asked to press the space bar while fixating on the dot. This was to check calibration at the beginning of each trial and also to direct participants' eye to the sentence-initial position. With acceptable calibration, pressing the space bar replaced the dot with a sentence, which was presented in a single line. Participants were asked to read the sentences silently for comprehension at their own pace and press the space bar one more time when they finished. This prompted a true/false statement for about half of the trials, to which the participants responded by pressing the 'true' key (the 'F' key of the keyboard) or the 'false' key ('J'). The eye-tracker was recalibrated at pre-determined breaks (once every about 10th trial.) Recalibration was additional conducted in case a participant was not able to see a sentence due to poor calibration accuracy. At the experimental setting, about 3.5 characters subtended 1 degree of visual angle. Sampling rate was 1000Hz. Participants were allowed to take a short rest if they wanted, at one of the designated re-calibration times.

After the participants completed the eye-tracking task, they completed the other tasks in the order of the reading span test, the subtract-2 span test, the language background questionnaire and the cloze test.

5.1.5. Data analysis

The data were analyzed for four critical words—the verb creating the first potential gap site (i.e. *wrote* in (2)), the immediately following adverb (*passionately*), the actual subcategorizer (*about/saw*) and the immediately following word (*while*).

Four eye movement measures were computed: *First-pass reading time*, *first-pass regression*, *regression path duration* and *total reading time*. The *first-pass reading time* is the sum of all fixations in a region before the eye leaves the region for the first time, either to the left or to the right, and has often been taken to reflect an early stage of processing. The *first-pass regression* represents the probability that a regressive eye movement is made after a region is fixated for the first time. The measure reflects a difficulty in integrating a word when it is first fixated, which is arguably an early effect (Clifton, Staub & Rayner, 2007). The *regression path duration* refers to the sum of all fixations from first fixating a region until the eye leaves the region to the right, including the time spent rereading the previous words. The regression path duration reflects a difficulty in integrating a newly encountered word and also the cost of overcoming this difficulty. It therefore could be taken to reflect a later stage of processing than the first-pass reading time or the first-pass regression. All the three measures discussed so far were calculated from only the trials in which the given region was not skipped for first pass reading. The *total reading time* is the sum of all fixations on a region, including any rereading times, regardless of whether the region was skipped for first pass reading. The total reading time therefore is a measure that reflects a later stage of processing compared to the other three measures.

Prior to the analysis, fixations less than 80ms within one degree of visual arc of another fixation was merged with that fixation. Among the remaining fixations, those shorter than 80ms or longer than 1000ms were removed.

The three reading times (first-pass reading time, regression path duration and total reading time) were statistically analyzed for each group using linear mixed effects modeling. For the first-pass regression, which is a categorical dependent variable, mixed effects logit modeling was conducted instead. In all models, the main effects and interactions of the two experimentally manipulated factors (i.e., Plausibility [Plausible vs. Implausible] and Island [Non-island vs. Island]) and the participants' working memory capacity (the composite working memory scores) were included as the fixed effects. The composite working memory scores were centered separately for each group, prior to being included in the models. The other available participant variables (cloze test score, age of first instruction, age of first immersion, immersion duration and percent English use) were not included as the fixed effects because preliminary analyses revealed that they do not make significant contributions to the models. Participant and item intercepts were included in the random effect structure. Plausibility and Island were coded using mean-centered contrast coding. For the linear mixed effect models *p*-values for the fixed effects were calculated through parameter estimation based on Markov chain Monte Carlo (mcmc) sampling.

5.1.6. Predictions

The predictions are basically similar to those of Experiment 1. Minor differences are discussed below.

- Unlike in Experiment 1, in Experiment 2 each critical word will be analyzed separately, rather than pre-set regions. The effects related to the active gap search will occur at the verb, although the effect might also spill over to following words. The 'reanalysis effect', if it occurs, could occur at the actual subcategorizer, but it is more likely to occur at the immediately following word (*while*) as shown in Experiment 1.
- The dependent measures are reading times derived from eye fixations (first-pass reading time, regression path duration and total reading time) and the first-pass regression, rather than the stop responses and reaction times. In Experiment 2, the conditions that are predicted to cause relative processing difficulty will induce longer reading times and/or more first-pass regressions.

5.2. Results and discussion

5.2.1. Comprehension accuracy

Response to the true/false statements for the experimental items was examined prior to conducting the analysis of the eye movement data. Native speakers' average comprehension accuracy was 89.6%, with a range from 64.3% to 100%. (SD=11.5). Among the L2 learners, one participants' comprehension accuracy score was less than chance (43.7%). Therefore, the participant's data were excluded from the main analysis. The average comprehension accuracy of the remaining L2 learners was 84% (Range=65%~100%, SD=10.2).

5.2.2. Eye movements of native speakers

Table 5.3 presents a descriptive summary of the eye movement measures (reading times and first-pass regression) of the native speaker group for each critical word.

Table 5.3. Means and standard deviations of the eye movement measures of the critical words: Native speakers

	first possible	first possible	actual	actual
First-pass reading time				
Non-island/plausible	289 (109)	300 (152)	235 (93)	290 (134)
Non-island/implausible	309 (164)	310 (147)	260 (90)	247 (113)
Island/plausible	276 (100)	298 (167)	310 (148)	289 (142)
Island/implausible	288 (156)	286 (124)	310 (133)	290 (156)
First-pass regression				
Non-island/plausible	0.14 (0.35)	0.23 (0.43)	0.07 (0.25)	0.44 (0.50)
Non-island/implausible	0.13 (0.33)	0.32 (0.47)	0.25 (0.43)	0.39 (0.49)
Island/plausible	0.31 (0.47)	0.21 (0.41)	0.13 (0.34)	0.24 (0.43)
Island/implausible	0.33 (0.47)	0.24 (0.43)	0.19 (0.39)	0.22 (0.42)
Regression path duration				
Non-island/plausible	413 (624)	477 (501)	278 (247)	522 (647)
Non-island/implausible	417 (357)	564 (508)	494 (711)	488 (489)
Island/plausible	476 (641)	527 (691)	370 (212)	525 (816)
Island/implausible	433 (334)	471 (460)	494 (535)	565 (867)
Total reading time				
Non-island/plausible	539 (395)	632 (413)	459 (310)	604 (391)
Non-island/implausible	598 (381)	656 (385)	453 (277)	495 (283)
Island/plausible	685 (456)	744 (457)	698 (464)	662 (466)
Island/implausible	571 (445)	601 (440)	603 (455)	561 (426)

Tables 5.4 – 5.7 show the fixed effects of the models fitted to the measures obtained from each word. The results of the statistical analyses will be presented for each critical word.

Table 5.4. fixed effects of the model fitted to the measures of native speakers' eye movement at the first possible subcategorizer (*wrote*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	290.098	12.525	23.162	0.000	***
	Plausibility	14.177	15.377	0.922	0.356	
	Island	-18.408	15.421	-1.194	0.242	
	WM	0.356	1.314	0.271	0.763	
	Plausibility:Island	-8.619	30.712	-0.281	0.776	
	Plausibility:WM	-0.180	1.956	-0.092	0.951	
	Island:WM	2.156	1.924	1.121	0.284	
	Plausibility:Island:WM	0.567	3.858	0.147	0.881	
First-pass regression	(Intercept)	-1.371	0.178	-7.723	0.000	***
	Plausibility	-0.005	0.301	-0.016	0.987	
	Island	1.218	0.301	4.041	0.000	***
	WM	-0.031	0.020	-1.559	0.119	
	Plausibility:Island	0.196	0.601	0.326	0.744	
	Plausibility:WM	-0.025	0.035	-0.722	0.470	
	Island:WM	0.016	0.035	0.452	0.651	
	Plausibility:Island:WM	-0.039	0.069	-0.56	0.575	
Regression path duration	(Intercept)	435.336	38.185	11.401	0.000	
	Plausibility	-19.816	56.518	-0.351	0.726	
	Island	35.706	56.736	0.629	0.530	
	WM	-1.315	4.728	-0.278	0.781	
	Plausibility:Island	-46.578	112.976	-0.412	0.680	
	Plausibility:WM	-3.182	6.998	-0.455	0.650	
	Island:WM	3.067	6.993	0.439	0.661	
	Plausibility:Island:WM	-6.904	13.982	-0.494	0.622	
Total reading time	(Intercept)	596.317	40.185	14.839	0.000	***
	Plausibility	-30.706	41.557	-0.739	0.461	
	Island	58.662	41.562	1.411	0.159	
	WM	-3.74	4.345	-0.861	0.390	
	Plausibility:Island	-182.385	83.113	-2.194	0.029	*
	Plausibility:WM	-4.623	5.359	-0.863	0.389	
	Island:WM	6.823	5.253	1.299	0.195	
	Plausibility:Island:WM	5.295	10.492	0.505	0.614	

* *t* values are reported for the reading times and *z* values for the first-pass regressions

* The number of asterisks besides the p values indicates the level of significance at 0.001***, 0.01**, 0.05*, 0.1†

At the first available subcategorizer (*wrote*), there was a significant main effect of island in the first-pass regression, with more first-pass regressions in the Island conditions than in the Non-island conditions. In total reading times, there was a significant interaction between Plausibility and Island, due to the different directions of the plausibility effect in the two levels of the Island conditions. As Figure 5.1 below shows, which presents the mean total reading times in the four conditions, the reading time was 59ms longer in the Implausible than the Plausible condition in the Non-island conditions, whereas it was 114ms longer in the Plausible than the Implausible condition in the Island conditions.

Figure 5.1. The total reading times in the four conditions at the first available subcategorizer

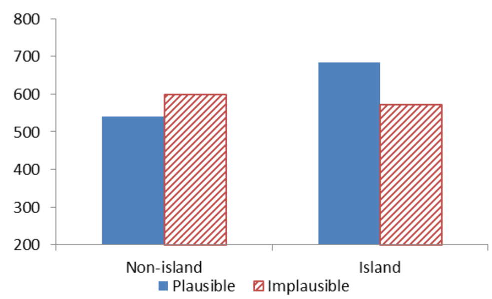


Table 5.5. Fixed effects of the model fitted to the measures of native speakers' eye movement at the word immediately following the first possible subcategorizer (*passionately*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	297.153	17.146	17.331	0.000	***
	Plausibility	0.447	13.906	0.032	0.984	
	Island	-15.495	13.898	-1.115	0.28	
	WM	1.404	1.820	0.771	0.421	
	Plausibility:Island	-21.412	27.809	-0.770	0.464	
	Plausibility:WM	-0.181	1.821	-0.099	0.896	
	Island:WM	0.854	1.769	0.483	0.595	
	Plausibility:Island:WM	-4.851	3.555	-1.364	0.185	
First-pass regression	(Intercept)	-1.159	0.156	-7.446	0.000	***
	Plausibility	0.288	0.248	1.159	0.247	
	Island	-0.311	0.248	-1.253	0.21	
	WM	-0.014	0.019	-0.707	0.48	
	Plausibility:Island	-0.301	0.497	-0.606	0.544	
	Plausibility:WM	-0.046	0.031	-1.476	0.14	
	Island:WM	-0.010	0.031	-0.312	0.755	
	Plausibility:Island:WM	-0.023	0.063	-0.375	0.708	
Regression path duration	(Intercept)	507.988	35.151	14.451	0.000	***
	Plausibility	16.499	56.084	0.294	0.769	
	Island	-23.794	56.085	-0.424	0.672	
	WM	2.406	4.385	0.549	0.584	
	Plausibility:Island	-138.609	112.201	-1.235	0.218	
	Plausibility:WM	-14.541	7.021	-2.071	0.039	*
	Island:WM	-0.595	7.021	-0.085	0.933	
	Plausibility:Island:WM	-24.468	14.053	-1.741	0.083	†
Total reading time	(Intercept)	657.353	46.870	14.025	0.000	
	Plausibility	-59.046	39.675	-1.488	0.138	
	Island	25.887	39.674	0.653	0.515	
	WM	-5.017	5.190	-0.967	0.334	
	Plausibility:Island	-167.364	79.36	-2.109	0.036	*
	Plausibility:WM	-1.541	5.219	-0.295	0.768	
	Island:WM	4.267	5.074	0.841	0.401	
	Plausibility:Island:WM	3.086	10.190	0.303	0.762	

At the adverb immediately following the first subcategorizer (*passionately*), there was a marginally significant interaction between Plausibility, Island and Working Memory as well as a significant interaction between Plausibility and Working Memory in the regression path duration. In order to understand the source of the trend of three-way interactions, the native speakers were divided into two groups based on their composite working memory scores (10 whose scores were equal to or higher than the median and 9 whose scores were lower than the median), and the regression path duration was examined in each group. Figure 5.2 shows the mean regression path duration of the four conditions in each group.

Figure 5.2. Regression path durations (a) in the higher working memory group (a) and in the lower working memory group (b) at the word immediately following the first subcategorizer

(a) higher working memory group (b) lower working memory group

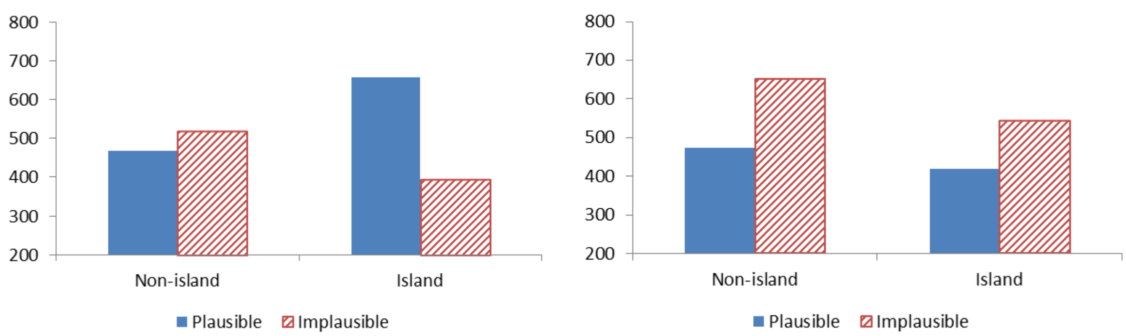


Figure 5.2 suggests that the higher working memory group show an interaction between Plausibility and Island, with different directions of plausibility effects in the two Island conditions. In the lower working memory group, however, the direction of plausibility effects were the same in the two Island conditions: The reading times in the Implausible conditions were longer than those of the Plausible conditions in both the Non-island and the Island conditions.

In the total reading time, there was a significant interaction between Plausibility and Island, which was not modulated by Working Memory. Examination of the four conditions revealed that in the Non-island conditions, the total reading time was 24ms longer in the Implausible than in the Plausible condition whereas in the Island conditions, it was 143ms longer in the Plausible than in the Implausible condition, a pattern similar to one that occurred in the total reading time at the previous region.

Figure 5.3. The total reading times in the four conditions at the adverb

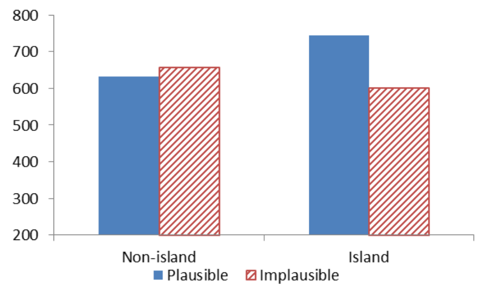


Table 5.6. Fixed effects of the model fitted to the measures of native speakers' eye movement at the actual subcategorizer (*about/saw*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	277.694	9.973	27.846	0.0001	***
	Plausibility	12.668	15.873	0.798	0.411	
	Island	62.354	15.911	3.919	0.0001	***
	WM	-1.580	1.324	-1.193	0.202	
	Plausibility:Island	-27.243	31.644	-0.861	0.407	
	Plausibility:WM	-0.045	2.117	-0.021	0.974	
	Island:WM	0.474	2.142	0.221	0.788	
	Plausibility:Island:WM	-3.357	4.220	-0.795	0.445	
First-pass regression	(Intercept)	-1.962	0.270	-7.278	3.39E-13	***
	Plausibility	1.057	0.439	2.409	0.016	*
	Island	0.145	0.441	0.328	0.743	
	WM	-0.015	0.027	-0.542	0.588	
	Plausibility:Island	-1.238	0.876	-1.413	0.158	
	Plausibility:WM	0.052	0.055	0.952	0.341	
	Island:WM	0.051	0.056	0.916	0.36	
	Plausibility:Island:WM	0.0541	0.112	0.485	0.628	
Regression path duration	(Intercept)	409.503	39.867	10.272	0.000	***
	Plausibility	175.337	60.967	2.876	0.004	**
	Island	39.569	61.364	0.645	0.520	
	WM	-0.527	4.515	-0.117	0.907	
	Plausibility:Island	-91.924	121.581	-0.756	0.450	
	Plausibility:WM	8.827	8.170	1.08	0.281	
	Island:WM	11.613	8.253	1.407	0.161	
	Plausibility:Island:WM	-1.774	16.436	-0.108	0.914	
Total reading time	(Intercept)	487.276	47.907	10.171	0.000	***
	Plausibility	-46.63	36.71	-1.27	0.205	
	Island	232.066	36.708	6.322	0.000	***
	WM	-5.447	5.344	-1.019	0.309	
	Plausibility:Island	-91.483	73.416	-1.246	0.214	
	Plausibility:WM	7.765	4.761	1.631	0.104	
	Island:WM	4.536	4.613	0.983	0.326	
	Plausibility:Island:WM	-4.151	9.226	-0.45	0.653	

At the actual subcategorizer (*about/saw*), the main effect of Island was significant in both the first-pass reading time and in the total reading time due to longer reading times in the Island conditions than the Non-island conditions.

In the first-pass regression and regression path duration, however, there was only a significant main effect of Plausibility, which was not modulated by Island, due to the Implausible conditions having more regressions and longer reading times than the Plausible conditions. Figure 5.4 show the mean first-pass regression and regression path duration in the four conditions.

Figure 5.4. Mean first-pass regression (a) and regression path duration (b) at the actual subcategorizer

(a) first-pass regression (b) regression path duration

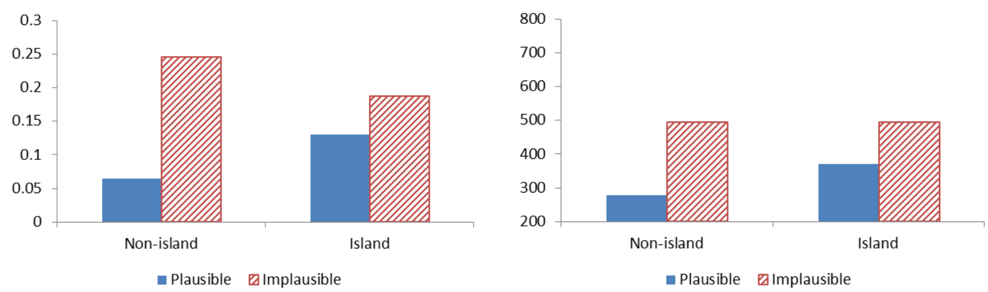


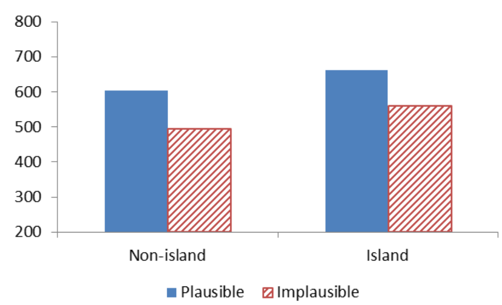
Table 5.7. Fixed effects of the model fitted to the measures of native speakers' eye movement at the word immediately following the actual subcategorizer (*while*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	-19.647	14.746	-1.332	0.0001	***
	Plausibility	24.082	14.815	1.626	0.186	
	Island	0.313	1.559	0.201	0.126	
	WM	45.037	29.551	1.524	0.818	
	Plausibility:Island	0.744	1.960	0.38	0.133	
	Plausibility:WM	1.882	1.907	0.987	0.712	
	Island:WM	3.164	3.888	0.814	0.358	
	Plausibility:Island:WM	-19.647	14.746	-1.332	0.455	
First-pass regression	(Intercept)	-0.840	0.216	-3.895	0.0001	***
	Plausibility	-0.177	0.269	-0.66	0.510	
	Island	-0.881	0.270	-3.267	0.001	**
	WM	-0.033	0.025	-1.321	0.187	
	Plausibility:Island	0.050	0.539	0.094	0.925	
	Plausibility:WM	0.019	0.035	0.538	0.591	
	Island:WM	-0.002	0.034	-0.058	0.953	
	Plausibility:Island:WM	0.012	0.070	0.179	0.858	
Regression path duration	(Intercept)	552.151	81.456	6.779	0.0000	***
	Plausibility	31.895	76.667	0.416	0.678	
	Island	72.593	76.842	0.945	0.346	
	WM	-3.366	4.886	-0.689	0.492	
	Plausibility:Island	81.624	152.834	0.534	0.594	
	Plausibility:WM	-7.973	10.344	-0.771	0.442	
	Island:WM	-6.813	9.951	-0.685	0.494	
	Plausibility:Island:WM	-10.462	20.159	-0.519	0.604	
Total reading time	(Intercept)	574.451	50.534	11.368	0.0000	***
	Plausibility	-109.461	36.795	-2.975	0.003	**
	Island	66.874	36.763	1.819	0.070	†
	WM	-2.032	5.017	-0.405	0.686	
	Plausibility:Island	3.18	73.488	0.043	0.966	
	Plausibility:WM	5.13	4.955	1.035	0.301	
	Island:WM	-1.386	4.805	-0.288	0.773	
	Plausibility:Island:WM	6.622	9.581	0.691	0.490	

At the word immediately following the actual subcategorizer (*while*), the main effect of Island was significant in the first-pass regression, due to a higher number of trials with the first-pass regression in the Non-island than in the Island conditions. Marginal main effect of Island in the reverse direction was found in the total reading times.

There was also a significant main effect of Plausibility in the total reading time, due to longer reading times in the Plausible than in the Implausible conditions as shown in Figure 5.5 below.

Figure 5.5. The total reading time in the four conditions at the word immediately following the actual subcategorizer (*while*)



5.2.3. Discussion of the native speakers' eye movements

The native speakers' eye movement data suggest a more complicated relationship between active gap search and the subject/relative clause island than in the stop-making-sense experiment in Experiment 1. It appears to be helpful to interpret the results in terms of the 'latest' measure (total reading time), and the other measures (first-pass reading time, first-pass regression and regression path duration), which tap into earlier processing compared to the total reading time.

Among the effects found in the measures of earlier processing, two are relevant to the issue of how native speakers use island constraints for filler-gap processing. At the word immediately following the first subcategorizer (*passionately*), there was a trend of interaction between Plausibility, Island and Working Memory in the regression path duration. This was because the higher span native speakers showed a trend of interaction between Plausibility and Island (i.e., evidence for their sensitivity to island constraints) whereas the lower span native speakers showed a trend of a Plausibility main effect with longer reading times in the Implausible conditions (i.e., evidence for evaluating a gap regardless of whether there is an island or not). At the next word (*about*/*saw*), there was a Plausibility main effect in the first-pass regression and regression path duration, with more regressions/longer reading times in the Implausible conditions.

These results have the following implications. First, the trend of three-way interaction at the word immediately following the first subcategorizer suggests that individual working memory capacity might be relevant to sensitivity to island constraints. Interestingly, the pattern of the interaction was the opposite of what the processing-based accounts of island constraints predict. The processing-based accounts predict that individuals with greater working memory capacity are more likely to posit a gap even within an island than those with smaller capacity because it is less likely to be demanding for them to reactivate and integrate a gap within a taxing structure such as an island. In this study, however, it was those with smaller capacity who posited a gap within an island, which is not expected if island insensitivity is a consequence of having enough working memory capacity to handle multiple demanding computations.

Unlike the processing-based accounts, previous grammar-based accounts of islands have not so far made a claim that individual working memory capacity is correlated with an island effect (Sprouse et al., 2012). However, the pattern of the relationship between the island effect and working memory capacity shown in this experiment is not unexpected under the assumption that island constraints are encoded in grammar and that a large working memory capacity offers an advantage in incorporating various types of information, including grammatical information, rapidly and efficiently in real time processing. The default strategy to process sentences with a filler is to posit a gap at the earliest possible gap position. In case the earliest possible gap position is within an island, the parser that is immediately sensitive to grammar must hold the filler in working memory until it finishes processing the island domain and detects the actual gap position outside the island. It is reasonable to hypothesize that this taxes working memory and is better performed by individuals with a large working memory capacity. Those who do not have enough working memory capacity to keep holding the filler in working memory until the island domain has been processed might end up reactivating the filler (i.e., form a filler-gap dependency) within an island, delaying evaluation of whether or not the dependency is grammatically licit. Note that this hypothesis shares an underlying assumption that an island is difficult to process with the processing-based accounts of islands of Kluender and colleagues (Kluender, 1998, 2004; Kluender & Kutas, 1993) and Hofmeister and Sag (2010). However, unlike these processing-based accounts, this hypothesis is based on the assumption that the island constraints have been grammaticalized. In this hypothesis, island constraints are something that must be followed if working memory capacity allows, unlike in the processing-based accounts in which they are something that can be overcome with enough working memory capacity.

Second, the fact that there was a main effect Plausibility at the word following the first subcategorizer (for the lower span native speakers) and at the next word (for the entire native speakers) suggests that even in native speakers' processing, evaluating an ungrammatical filler-gap dependency may not be completely prevented by island constraints. This finding is unexpected given that Experiment 1 and many other previous L1 processing studies showed that native speakers are immediately sensitive to island constraints. However, it needs to be mentioned that this study is not the only one that reports that the native speaker participants were insensitive to islands in real time processing (Clifton & Frazier, 1989; Freedman & Forster, 1985; Pickering et al., 1994). Therefore, the possibility will be taken into consideration that the result of this study is a valid piece of data suggesting that at least under certain experimental circumstances native speakers may not be able to completely suppress active gap search within an island.

In the total reading time, which reflects a very late stage of processing (as well as earlier stages of processing), there was an interaction between Plausibility and Island, without further interaction with Working Memory, at the first subcategorizer and at the next word (*wrote passionately*). There was also a main effect of Plausibility at the word following the actual subcategorizer (*while*), due to longer reading times in the Plausible conditions than the Implausible conditions. If we examine the patterns of these effects, they are similar to those predicted if the parser is sensitive to both the grammatical constraint (i.e., island constraints) and the plausibility information (i.e., plausibility of the wh-dependency to be formed) in determining whether to form a wh-dependency (see Prediction 3 in Chapter 4). The native speakers, regardless of their individual working memory capacity, were ultimately sensitive to island constraints, but the grammar does not appear to have been the only constraint affecting their parsing decisions. For some reason, the plausibility of the filler as an argument of the first subcategorizer seems to have a continuing effect in parallel, causing the grammatical constraint to have a less deterministic effect in completely ruling out an ungrammatical (but plausible) analysis even at a very late stage. This is not consistent with the results of Experiment 1, in which the native speakers' stop responses and reading times suggest that the effect of island constraints had overridden the effect of plausibility (Prediction 2 in Chapter 4). Assuming the differences in lexical items of the experimental sentences of the two experiments had a minor influence, one factor that is likely to have caused the difference is the task given to the participants. In a stop-making-sense task, participants read sentences one region at a time and are not able to go back to the previous regions because the previous regions are not presented. The participants also have to monitor the plausibility of the sentences at each segment. An eye-tracking task provides an environment that is more similar to typical reading situations than in the stop-making-sense task. Therefore, an eye-tracking task is less likely to force participants to read carefully and filter out ungrammatical representations than a stop-making-sense task. Although the native speakers in Experiment 2 do not show evidence of committing only to the grammatically accurate parse, the results show that the grammatical constraint is at least one of the constraints guiding native speakers' (quasi-)typical reading. Whether L2 learners' reading is also guided by island constraints is examined in the next section.

5.2.4. Eye movements of L2 learners

A descriptive summary of the eye movements of L2 learners are provided in Table 5.8 for each critical word.

Table 5.8. Means and standard deviations of the eye movement measures of the critical words: L2 learners

	first possible	first possible	actual	actual
First-pass reading time				
Non-island/plausible	383 (197)	538 (351)	292 (131)	371 (169)
Non-island/implausible	384 (192)	588 (370)	324 (140)	370 (235)
Island/plausible	338 (153)	511 (275)	368 (151)	456 (286)
Island/implausible	350 (174)	531 (290)	374 (191)	394 (189)
First-pass regression				
Non-island/plausible	0.19 (0.39)	0.19 (0.40)	0.12 (0.33)	0.18 (0.38)
Non-island/implausible	0.21 (0.41)	0.21 (0.41)	0.07 (0.25)	0.26 (0.44)
Island/plausible	0.20 (0.40)	0.23 (0.42)	0.15 (0.36)	0.36 (0.48)
Island/implausible	0.27 (0.45)	0.15 (0.36)	0.16 (0.37)	0.33 (0.47)
Regression path duration				
Non-island/plausible	566 (504)	702 (453)	378 (331)	465 (283)
Non-island/implausible	601 (664)	824 (617)	349 (177)	557 (601)
Island/plausible	519 (603)	794 (748)	608 (887)	889 (930)
Island/implausible	669 (935)	771 (799)	572 (705)	781 (875)
Total reading time				
Non-island/plausible	711 (543)	909 (596)	524 (451)	690 (515)
Non-island/implausible	755 (552)	1112 (769)	518 (327)	768 (754)
Island/plausible	833 (566)	1247 (1006)	970 (726)	983 (664)
Island/implausible	851 (617)	1134 (794)	945 (760)	939 (698)

The fixed effects of the models fitted to the eye movement measures at the critical words are presented in Tables 5.9 – 5.12 below.

Table 5.9. Fixed effects of the model fitted to the measures of L2 learners' eye movement at the first possible subcategorizer (*wrote*)

		Estimate	Std. error	*t*/*z* value	*p* value	
First-pass reading time	(Intercept)	362.938	20.089	18.066	0.0001	***
	Plausibility	9.225	15.741	0.586	0.567	
	Island	-38.937	15.754	-2.472	0.014	*
	WM	0.504	1.608	0.314	0.726	
	Plausibility:Island	13.549	31.537	0.43	0.672	
	Plausibility:WM	1.3285	1.489	0.892	0.379	
	Island:WM	0.993	1.473	0.674	0.511	
	Plausibility:Island:WM	3.655	2.946	1.24	0.223	
First-pass regression	(Intercept)	-1.376	0.152	-9.073	0.0000	***
	Plausibility	0.250	0.251	0.995	0.320	
	Island	0.221	0.251	0.883	0.377	
	WM	-0.027	0.014	-1.932	0.053	†
	Plausibility:Island	0.242	0.502	0.482	0.630	
	Plausibility:WM	-0.037	0.023	-1.622	0.105	
	Island:WM	0.034	0.023	1.475	0.140	
	Plausibility:Island:WM	-0.052	0.046	-1.146	0.252	
Regression path duration	(Intercept)	589.762	50.276	11.731	0.0000	***
	Plausibility	92.824	65.638	1.414	0.158	
	Island	9.467	65.612	0.144	0.885	
	WM	-3.946	4.706	-0.839	0.402	
	Plausibility:Island	117.096	131.264	0.892	0.373	
	Plausibility:WM	-5.716	6.122	-0.934	0.351	
	Island:WM	13.404	6.119	2.191	0.029	*
	Plausibility:Island:WM	2.691	12.238	0.22	0.826	
Total reading time	(Intercept)	790.932	80.737	9.796	0.0000	***
	Plausibility	38.01	43.452	0.875	0.382	
	Island	116.383	43.512	2.675	0.008	**
	WM	5.264	7.158	0.735	0.463	
	Plausibility:Island	-1.829	87.025	-0.021	0.983	
	Plausibility:WM	-3.615	4.105	-0.881	0.379	
	Island:WM	7.291	4.071	1.791	0.074	†
	Plausibility:Island:WM	-2.145	8.141	-0.263	0.792	

At the first subcategorizer (*wrote*), the main effect of Island was significant in the first-pass reading time due to longer reading times in the Non-island conditions than in the Island conditions, which was the opposite of what was found in the native speaker group at the same region in the first-pass regression (more regressions in the Island than the Non-island conditions, see Table 5.4 above). The effect of the Island was modulated by Working Memory in the regression path duration due to the fact that the learners with the smaller working memory capacity (11 learners whose composite working memory scores were lower than the median) read the Non-island conditions slower than the Island conditions (Non-island: 681ms vs. Island: 593ms) whereas the learners with the higher working memory capacity (11 learners whose composite working memory scores were higher than the median) read the Island conditions slower (Non-island: 486ms, Island: 593ms). In the total reading time, the main effect of Island was significant, with the reading time being longer in the Island conditions than in the Non-island conditions. The marginal interaction between Island and Working Memory reflects the tendency that the difference between the two conditions was larger in the higher working memory group (Non-island: 765ms, Island: 920ms) than in the lower working memory group (Non-island: 700ms, Island: 765ms). Additionally, the main effect of Working Memory was close to significant in the first-pass regression, due to the tendency for the lower span learners to make more first-pass regressions (26%) than the higher span learners (16%).

Table 5.10. Fixed effects of the model fitted to the measures of L2 learners' eye movement at the word immediately following the first possible subcategorizer (*passionately*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	541.418	39.472	13.716	0.0001	***
	Plausibility	39.128	26.830	1.458	0.147	
	Island	-40.902	26.875	-1.522	0.124	
	WM	1.541	2.722	0.566	0.527	
	Plausibility:Island	-26.769	53.735	-0.498	0.629	
	Plausibility:WM	-2.982	2.546	-1.171	0.256	
	Island:WM	-1.452	2.519	-0.576	0.554	
	Plausibility:Island:WM	0.650	5.036	0.129	0.885	
First-pass regression	(Intercept)	-1.502	0.163	-9.232	0.000	***
	Plausibility	-0.212	0.252	-0.842	0.4	
	Island	-0.116	0.252	-0.459	0.646	
	WM	0.011	0.014	0.824	0.41	
	Plausibility:Island	-0.753	0.505	-1.492	0.136	
	Plausibility:WM	-0.029	0.024	-1.238	0.216	
	Island:WM	-0.026	0.023	-1.097	0.273	
	Plausibility:Island:WM	-0.012	0.047	-0.263	0.792	
Regression path duration	(Intercept)	772.335	43.196	17.88	0.0000	***
	Plausibility	51.427	62.959	0.817	0.415	
	Island	19.587	62.959	0.311	0.756	
	WM	2.885	4.04	0.714	0.476	
	Plausibility:Island	-143.099	125.919	-1.136	0.256	
	Plausibility:WM	-10.144	5.873	-1.727	0.085	†
	Island:WM	-4.221	5.873	-0.719	0.473	
	Plausibility:Island:WM	-8.694	11.745	-0.74	0.460	
Total reading time	(Intercept)	1101.517	110.091	10.005	0.0000	***
	Plausibility	58.619	61.877	0.947	0.344	
	Island	177.146	61.979	2.858	0.005	**
	WM	9.604	9.240	1.039	0.299	
	Plausibility:Island	-307.997	123.898	-2.486	0.013	*
	Plausibility:WM	-11.05	5.872	-1.882	0.061	†
	Island:WM	-0.829	5.817	-0.142	0.887	
	Plausibility:Island:WM	-16.326	11.627	-1.404	0.161	

At the word immediately following the first subcategorizer (*passionately*), the interaction between Plausibility and Working Memory was close to significant in the regression path duration and in the total reading time. This was because the lower working memory group took longer to read the Implausible conditions than the Plausible conditions ([regression path duration] Plausible: 666ms, Implausible: 841ms; [total reading time] Plausible: 922ms, Implausible: 1082ms), whereas the higher working memory group took longer to read the Plausible conditions than the Implausible conditions ([regression path duration] Plausible: 828ms, Implausible: 754ms; [total reading time] Plausible: 1234ms, Implausible: 1164ms). The direction of the Plausibility effect in the higher working memory group does not conform to any of the three predictions made earlier, and its implication to the issue of island sensitivity is not clear. Therefore, the data were examined more closely by calculating the regression path duration and total reading time of the four experimentally manipulated conditions within each working memory group, as shown in Figures 5.6 and 5.7, for exploratory purposes, even though the three-way interaction did not reach significance in either measure.

Figure 5.6. Regression path durations (a) in the higher working memory group (a) and in the lower working memory group (b) at the word immediately following the first subcategorizer (*passionately*)

(a) higher working memory group (b) lower working memory group

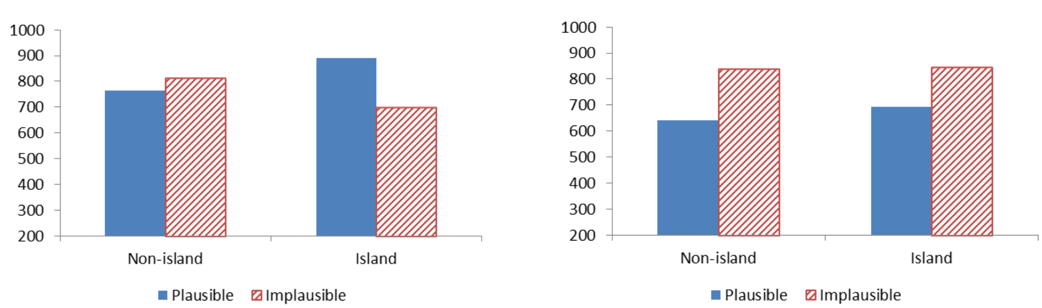
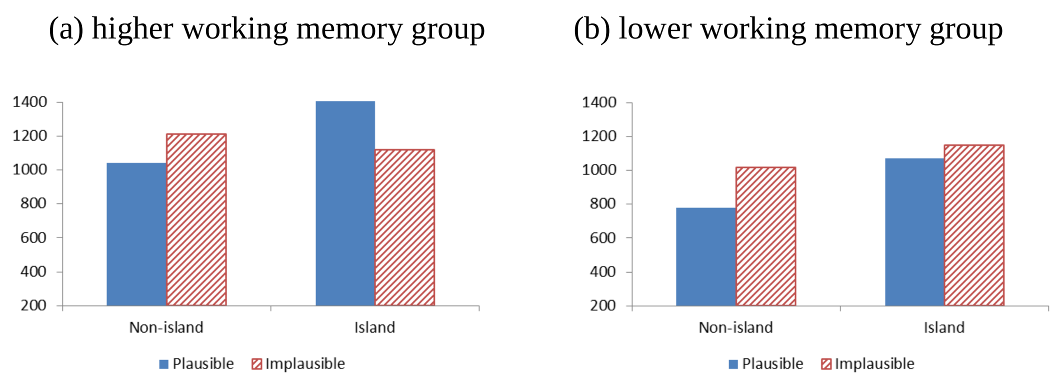


Figure 5.7. Total reading times (a) in the higher working memory group (a) and in the lower working memory group (b) at the word immediately following the first subcategorizer (*passionately*)



Figures 5.6 show that the numerically higher regression path duration of the Plausible conditions in the higher working memory group is mainly driven by the Island conditions. Interestingly, the numerical pattern of the regression path duration in the two groups shown in Figure 5.6 is strikingly similar to the pattern of the native speakers in the same measure at the same region (the regression path duration at the word following the first subcategorizer, see Figure 5.2), where the three-way interaction almost reached significance. A similar contrast between the two groups is revealed in the total reading time in Figure 5.7, although the lower working memory group seems to show a slightly smaller plausibility effect in the Island conditions compared to in the regression path duration.

In the total reading time, the interaction between Plausibility and Island reached significance (without interaction with Working Memory), as the following Figure 5.8 shows. Additionally, the main effect of Island was significant, due to a longer reading time of the Island conditions (Non-island: 1009ms, Island: 1190ms).

Figure 5.8. The total reading times in the four conditions at the word immediately following the first subcategorizer

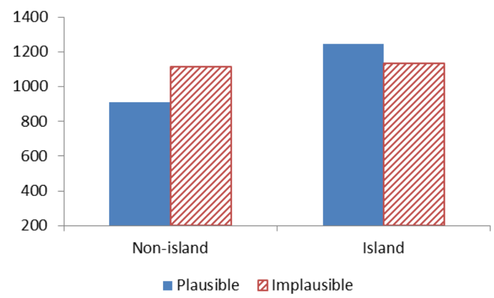


Table 5.11. Fixed effects of the model fitted to the measures of L2 learners' eye movement at the actual subcategorizer (*about/saw*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	338.348	17.331	19.523	0.0001	***
	Plausibility	22.734	15.137	1.502	0.141	
	Island	64.277	15.271	4.209	0.0001	***
	WM	-0.193	1.430	-0.135	0.876	
	Plausibility:Island	-25.678	30.277	-0.848	0.389	
	Plausibility:WM	1.018	1.431	0.711	0.501	
	Island:WM	0.544	1.423	0.382	0.714	
	Plausibility:Island:WM	3.659	2.850	1.284	0.195	
First-pass regression	(Intercept)	-2.430	0.308	-7.895	0.000	***
	Plausibility	-0.338	0.440	-0.769	0.442	
	Island	0.840	0.442	1.899	0.058	†
	WM	0.040	0.028	1.435	0.151	
	Plausibility:Island	0.916	0.879	1.042	0.297	
	Plausibility:WM	0.022	0.044	0.5	0.617	
	Island:WM	-0.075	0.045	-1.668	0.095	†
	Plausibility:Island:WM	-0.010	0.088	-0.117	0.907	
Regression path duration	(Intercept)	481.644	62.364	7.723	0.000	***
	Plausibility	-27.199	62.606	-0.434	0.664	
	Island	213.605	63.029	3.389	0.001	***
	WM	-0.168	5.217	-0.032	0.974	
	Plausibility:Island	-3.8	125.213	-0.03	0.976	
	Plausibility:WM	1.124	5.912	0.19	0.849	
	Island:WM	-3.788	5.885	-0.644	0.520	
	Plausibility:Island:WM	7.833	11.782	0.665	0.507	
Total reading time	(Intercept)	740.002	73.009	10.136	0.000	***
	Plausibility	-4.895	54.582	-0.09	0.929	
	Island	433.8	54.756	7.922	0.000	***
	WM	3.178	6.194	0.513	0.608	
	Plausibility:Island	-16.303	109.252	-0.149	0.882	
	Plausibility:WM	1.318	5.205	0.253	0.800	
	Island:WM	3.635	5.18	0.702	0.483	
	Plausibility:Island:WM	10.361	10.378	0.998	0.319	

At the actual subcategorizer (*about/saw*), the main effect of Island was significant or close to significant in all the four eye movement measures, due to longer reading times/more regressions in the Island conditions than in the Non-island conditions ([first-pass reading time] Non-island: 301ms, Island: 371ms; [first-pass regression] Non-island: 0.09, Island: 0.15; [regression path duration]: Non-island: 364ms, Island: 590ms; [total reading time] Nonisland: 521ms, Island: 958ms). In the first-pass regression, there was a tendency for the Island effect to be modulated by Working Memory, due to the fact that the Island effect was mainly driven by the lower working memory group ([higher WM]: Non-island: 0.15, Island: 0.19; [lower WM]: Non-island: 0.04, Island: 0.12).

Table 5.12. Fixed effects of the model fitted to the measures of L2 learners' eye movement at the word immediately following the actual subcategorizer (*while*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	391.895	26.391	14.85	0.0001	***
	Plausibility	-31.907	19.925	-1.601	0.111	
	Island	50.200	19.986	2.512	0.011	*
	WM	0.460	1.515	0.303	0.743	
	Plausibility:Island	-64.888	39.953	-1.624	0.109	
	Plausibility:WM	-0.871	1.904	-0.457	0.641	
	Island:WM	1.062	1.889	0.562	0.574	
	Plausibility:Island:WM	6.280	3.776	1.663	0.103	
First-pass regression	(Intercept)	-1.018	0.150	-6.782	0.000	***
	Plausibility	0.183	0.235	0.779	0.436	
	Island	0.652	0.235	2.773	0.006	**
	WM	0.012	0.013	0.877	0.381	
	Plausibility:Island	-0.688	0.471	-1.462	0.144	
	Plausibility:WM	0.016	0.022	0.696	0.486	
	Island:WM	0.039	0.022	1.761	0.078	†
	Plausibility:Island:WM	-0.029	0.045	-0.644	0.520	
Regression path duration	(Intercept)	673.208	46.857	14.367	0.000	***
	Plausibility	-11.445	70.839	-0.162	0.872	
	Island	326.165	70.971	4.596	0.000	***
	WM	3.88	3.337	1.163	0.246	
	Plausibility:Island	-202.964	141.875	-1.431	0.153	
	Plausibility:WM	-3.099	6.724	-0.461	0.645	
	Island:WM	15.621	6.698	2.332	0.020	
	Plausibility:Island:WM	4.005	13.389	0.299	0.765	
Total reading time	(Intercept)	828.805	89.373	9.274	0.000	***
	Plausibility	31.412	53.53	0.587	0.558	
	Island	224.643	53.61	4.19	0.000	***
	WM	5.933	6.377	0.93	0.353	
	Plausibility:Island	-122.53	107.217	-1.143	0.254	
	Plausibility:WM	2.684	5.131	0.523	0.601	
	Island:WM	11.256	5.083	2.214	0.027	*
	Plausibility:Island:WM	13.033	10.155	1.283	0.200	

At the word immediately following the actual subcategorizer, the main effect of Island was significant in all measures as in the previous region, again due to longer reading times/more regression in the Island conditions than in the Non-island conditions ([first-pass reading time] Non-island: 371ms, Island: 425ms; [first-pass regression] Non-island: 0.22, Island: 0.34; [regression path duration]: Non-island: 512ms, Island: 835ms; [total reading time] Nonisland: 730ms, Island: 961ms). The interaction between Island and Working Memory was marginally significant in the first-pass regression and was significant in the total reading time. This was because although both groups made more regressions and had longer total reading times in the Island than Non-island conditions, the difference was bigger in the higher working memory group ([higher WM/first-pass regression] Non-island: 0.22, Island: 0.38, [higher WM/total reading times] Non-island: 739ms, Island: 1051ms, [lower WM/first-pass regression] Non-island: 0.22, Island: 0.31, [lower WM/total reading times] Non-island: 721ms, Island: 872ms).

5.2.5. Discussion of L2 learners' eye movements

Sensitivity to island constraints

A noticeable difference between the native speakers' and the L2 learners' results is that the L2 learners were generally less sensitive to the plausibility manipulation. In the present experimental design, sensitivity to plausibility manipulation is important to see the role of island constraints in gap processing, but the effects related to the factor Plausibility were found only at the adverb following the first subcategorizer (*passionately*), in the form of a marginally significant interaction between Plausibility and Working Memory in the regression path duration/total reading time and a significant interaction between Plausibility and Island in the total reading time.

The results do not seem to provide clear answers with regard to the role of islands in the earlier stages of L2 filler-gap processing. It was suggestive, however, that for regression-path duration, the interaction of Plausibility and Working Memory came close to significance at the adverb. This interaction was due to the fact that the higher working memory learners spent more time rereading the earlier part of the sentence when they encountered the adverb in the Plausible than in the Implausible conditions while the lower working memory learners showed the opposite pattern. The result might suggest that the lower working memory learners tended to actively posit a gap regardless of whether the potential gap position was within an island or not, showing the typical plausibility effect. On the other hand, it is not obvious what the higher working memory learners' longer regression-path duration in the Plausible condition implies regarding their sensitivity to the island constraint. Further analyses on the regression-path duration within each working memory group provide some hints for interpreting this unexpected finding. As Figure 5.6 shows, the higher working memory learners' longer reading times in the Plausible conditions seem to be mainly due to the pattern of the reading times in the Island condition, with little notable plausibility effect in the non-island condition. The overall pattern seems to suggest that higher working memory learners distinguished between the Island and Non-island conditions in some way. The low working memory learners, on the other hand, took longer to read the Implausible sentences both in the Island and Non-island conditions, showing little sensitivity to the island constraints in filler-gap processing. Although the three-way interaction of Plausibility, Island and Working Memory did not reach significance, this trend of differential sensitivity to the island constraints between the learners with different working memory capacity is still interesting given that the native speakers showed a reliable three-way interaction of the same pattern at the same region and on the same measure as did the L2 learners.

It will be worthwhile to test a greater number of L2 participants in future research to examine whether this suggestive trend of the three-way interaction turns out to be reliable with greater power.

The total reading times showed a reliable interaction of Plausibility and Island, clearly suggesting that the L2 learners were sensitive to the presence of the island in processing filler-gap dependencies at a late stage of processing. The result suggests that L2 learners are able to apply grammatical constraints during online sentence processing even under normal reading situations, where they are not forced to interpret sentences in a strictly incremental manner as in Experiment 1.

Processing relative clauses

Experiment 2 also yielded some interesting results that might reflect the difference between L1 and L2 processing of relative clauses in general. Relevant to this issue is the differential patterns of the effects of Island that were not modulated by plausibility manipulation. Most interesting is that at the first subcategorizing verb (*wrote*), the direction of the main effects of Island was different between the native speakers and the L2 learners. The native speakers regressed more often in the Island conditions than in the Non-island conditions. In contrast, the L2 learners' first-pass reading times were longer in the Non-island conditions than in the Island conditions. The L2 learners' regression-path duration was modulated by their working memory capacity such that the lower working memory learners spent longer time reading the verb in the Non-island conditions than in the Island conditions, while the higher working memory learners showed the opposite pattern. In the total-reading times, the Island conditions caused longer reading times than the Non-island conditions for both groups of learners, with a numerically greater magnitude of the effect for the higher working memory learners.

If we assume that the main effects of Island are associated with the processing complexity involved in constructing the relative clause (*who wrote*) and the complex NP headed by the preceding noun (*author*), the results described above are consistent with the slower speed of the L2 learners in parsing and interpreting the target sentences as compared to the native speakers. The native speakers' greater processing difficulty as reflected in the increased frequency of first-pass regression at the relative clause verb (*wrote*) would suggest that they were able to construct the relative clause structure (*who wrote*) as part of the complex NP headed by the embedded clause subject (*the author*) as early enough as can be detected by their first-pass regression behaviors. On the other hand, the L2 learners' comparable first-pass reading times at the verbs in the Island and Non-island conditions would suggest that they failed to construct the relative clause in the Island conditions, thus showing no complexity effect, until they first encountered the relative clause verb. However, they came to show the native-like pattern of processing complexity in later measures, suggesting that it took a greater amount of time for them to construct the relative clause and the complex NP. The influence of learners' working memory capacity suggests that the high working memory learners might have been a little faster in the task of relative clause construction (as suggested by the finding that they showed the native-like pattern both in the regression-path duration and the total reading time) than the lower working memory learners. It is plausible that a larger working memory capacity might help L2 learners construct complex structures such as relative clauses and the associated syntactic relationships with the preceding heads.

At the actual subcategorizer and the following word, the patterns of the main effects of Island for the native speaker and the L2 learners converged such that it took more time to process the sentences in the Island than in the Non-island conditions. The greater processing complexity in the Island conditions in these regions could be attributed to the need to integrate the subject and the filler (i.e., the object) with the subcategorizing verb plus the difficulty processing the preceding complex subject NP. The fact that the L2 learners showed the native-like complexity effect seems to suggest that they were able to compute the preceding relative clause (and the complex NP) at this point in the sentences. Judging by the fact that the L2 learners showed a greater processing complexity effect in a greater number of measures than the native speakers, the processing of the complex NP seems to have been more demanding for the L2 learners than for the native speakers

Reanalysis

The native speakers showed sensitivity to plausibility manipulation at the actual subcategorizer and the following word, with longer total-reading times in the Plausible than in the Implausible condition. This pattern may have resulted from a reanalysis of the initial wh-dependency (see Prediction 3 in Chapter 4). Although it needs to be kept in mind that the present experiment did not directly probe the final resolutions of the filler-gap dependencies, this result suggests that the native speakers attempted to reanalyze the initially built wh-dependencies. The L2 learners, on the other hand, did not show any sensitivity to the plausibility manipulation at the actual subcategorizer and the following word, which is consistent with what was shown in Experiment 1. Taken together, the results suggest that L2 learners experience greater difficulty than native speakers in cancelling initially constructed wh-dependencies and forming new ones.

Chapter 6

Experiment 3

Experiments 1 and 2 investigated the way advanced Korean-speaking L2 learners of English process *wh*-dependencies, focusing on whether their *wh*-dependency processing is constrained by island constraints. Overall, the two experiments showed evidence that L2 learners are at least ultimately sensitive to island constraints in online *wh*-dependency formation, although application of the constraints may not be always automatized enough to prevent the parser from temporarily considering a grammatically-illegal, but resource-effective *wh*-dependency. Expanding the scope of the two studies, Experiment 3 investigated whether L2 filler-gap processing is also sensitive to constraints on parasitic gaps, which is a grammatical phenomenon interacting with island constraints.

Although islands are syntactic phrases in which a gap must not be posited, it has been reported that a gap in an island can be well-formed if it is 'parasitic' on another gap that is grammatically licit (Culicover & Postal, 2001, Engdahl, 1983). Examples in (1) illustrate the phenomenon (Engdahl, 1983, p.14).

(1) a. ?Here is the paper that John read his mail <before filing *t* >.

b. Here is the paper that John read *t* <before filing his mail>.

c. Here is the paper that John read *t* <before filing t_{pg} >.

In example (1), a temporal adjunct is an island in that a gap within a temporal adjunct degrades grammaticality (as in 1a), unlike a gap outside it (as in 1b). However, grammaticality improves if the illicit gap within the island occurs together with the licit gap outside the island and the two gaps share the antecedent (as in 1c). The gap that cannot be independent (as the second gap in 1c) has been called a 'parasitic gap' (hence the subscripted 'pg' in 1c).

It has been observed that parasitic gaps have a limited distribution and they can occur only within certain types of extraction islands. Engdahl (1983) proposed an accessibility hierarchy for occurrence of parasitic gaps as in (2). According to the hierarchy, manner adverbs are most likely to accommodate a parasitic gap whereas relative clauses/indirect questions are least likely to accommodate a parasitic gap (2a). In addition, parasitic gaps in untensed domains are more natural than those in tensed domains (2b).

(2) a. manner adverbs > temporal adverbs > purpose clauses > *that*/*than* clauses
> *when*/*because*/conditional *if* clauses > relative clauses/indirect questions

b. untensed domains > tensed domains

There is experimental evidence that parasitic gaps are restricted to a subclass of islands. In a grammaticality judgment study, Phillips (2006, Experiment 1) compared the grammaticality of sentences such as (3c) and (4c), in which a parasitic gap is placed within a subject NP with an infinitival complement and within a subject NP with a relative clause modifier, respectively. The results showed that (3c) is as good as (3b), suggesting that a parasitic gap can be placed naturally within a subject island that contains an infinitival complement. By contrast, (4c) was judged to be much worse than (4b). This shows that a subject island that contains a relative clause modifier does not readily allow a parasitic gap. Note that these results are compatible with Engdahl's accessibility hierarchy in that (3) involves a parasitic gap in an untensed domain whereas (4) involves a parasitic gap in a relative clause, which is a tensed domain.

(3) a. *What did [the attempt to repair *t*] ultimately damage the car?

b. What did [the attempt to repair the car] ultimately damage *t* ?

c. What did [the attempt to repair t_{pg}] ultimately damage *t* ?

(4) a. *What did [the reporter that criticized *t*] eventually praise the war?

b. What did [the reporter that criticized the war] eventually praise *t* ?

c. *What did [the reporter that criticized t_{pg}] eventually praise *t* ?

Phillips' Experiment 2 showed that the knowledge of the distribution of parasitic gaps influences online filler-gap processing as well. The self-paced reading study investigated whether native English speakers form a *wh*-dependency at the linearly first verb when processing sentence types such as (3) and (4) (at *repair* and *criticize*, respectively). Although the subject NPs in both (3) and (4) are islands for extraction, the subject NP in (3) supports a parasitic gap. Therefore, active gap creation in (3) may turn out to be correct, whereas active gap creation in (4) will always violate the grammar. Phillips found evidence for active gap creation in (3) but not in (4), and concluded that native language processing builds structures with substantial grammatical precision, by accurately and incrementally implementing the grammar of islands and parasitic gaps.

Having showed evidence that L2 learners are able to apply island constraints in processing of filler-gap dependency (although not always immediately) in the previous experiments, the present dissertation investigates whether L2 learners are sensitive to a more complicated constraint related to islands. Experiment 3 examined whether the learners are sensitive to the grammar of parasitic gaps, as well as the grammar of islands, which would indicate that the sentence structures they compute in real time comprehension are highly precise in terms of grammar. As in Experiment 2, Experiment 3 used the eye-tracking technique and examined whether working memory capacity influences the way readers use grammatical constraints in real-time processing.

6.1. Method

6.1.1. Participants

English proficiency and language background

Twenty-six native speakers of English (age 19-33, mean=23.2) and twenty-seven Korean-speaking L2 learners of English (age 20-38, mean=25) participated in Experiment 3. All participants were recruited from the community of University of Illinois at Urbana-Champaign. The L2 learners completed the same cloze test and the same language background questionnaire as were used in Experiment 1 and 2. The learners' cloze test scores and language backgrounds are summarized in Table 6.1. Table 6.1 shows the summary of the data received from 26 learners, after discarding the data from one learner whose eye-movement data were ultimately excluded from the main analysis due to poor comprehension accuracy (see below). For the purpose of comparison, the scores of Experiment 1 and 2 are also repeated in Table 6.1.

Table 6.1. L2 learners' cloze test scores and language background information in Experiment 1, 2 and 3

		cloze test	age of first	age of first	duration of	English use
Exp 1 (N=31)	range	28-37	7-14	15-34	Less than 1-8	5-100
	mean	33.6	11.1	23.1	3.6	40.8
	SD	2.4	1.9	5.1	2.4	26.8
Exp 2 (N=22)	range	27-38	6-15	15-31	Less than 1-9	5-70
	mean	32.8	11.2	21.6	5.1	38.1
	SD	3.1	2.1	5.7	2.8	18.2
Exp 3 (N=26)	range	28-37	6-16	14-33	Less than 1-8	5-85
	mean	33.1	10.2	21.2	3.4	40
	SD	2.5	2.7	5.5	2.5	20

As the table shows, all L2 learners who participated in Experiment 3 received the cloze test score of 28 or above out of 40, which is at or above 70% of the maximum score. Given that the structures of the experimental sentences were fairly complex, an advanced or close to advanced English proficiency was deemed necessary in order to comprehend the sentences. Although all learners received English instruction in Korea, some from an early age at six or seven, the age at which regular immersion in an English-speaking environment began was after they reached age 14 or above. Based on this background, the L2 learners can be roughly characterized as the advanced post-puberty learners. There were substantial variations in the duration of immersion and self-reported amount of English use per day.

Table 6.1 also shows that the L2 learner participants of the three experiments had closely comparable English proficiency and language background, except that the mean immersion duration of the participants of Experiment 2 was slightly higher than that of the other participants. Independent samples *t*-tests were conducted to check whether the L2 participants of Experiment 3 were indeed similar to the participants of Experiment 1 and 2 in each category. The participants of Experiment 3 were not significantly different from the participants of Experiment

1 in any category according to the independent samples *t*-tests (all *p*s>.1). When compared to the participants of Experiment 2, they had a significantly shorter duration of immersion experience ($p<.05$), but they were not different in the other categories (all *p*s>.1).

Working memory capacity

The two working memory tests used in Experiment 2 (reading span test and subtract-2-span test) were also administered to the native speakers and L2 learners under the same procedures. The same scoring methods were used. As for the reading span score, the accuracy of the native speakers' response on the plausibility judgment was 40 or above out of the highest possible score of 42 (95% or above), suggesting that all participants performed the processing component of the reading span task successfully. As was in Experiment 2, the L2 learners' plausibility judgment scores were lower than those of native speakers, with some quite low scores (range=22-42 (52.4%-100%), mean=36 (85.7%), SD=4.8 (11.3%)). However, no L2 participants were excluded from the main analysis as in Experiment 2.

A composite working memory score was calculated by translating the two test scores into percentage and averaging them. The score was used as the indicator of participants' working memory capacity. Table 6.2 presents a summary of the composite working memory score of the participants, together with their reading span score and the subtract-2 task score, which were included for reference. The corresponding scores of Experiment 2 are also repeated for ease of comparison.

Table 6.2. Native speakers' and L2 learners' working memory scores (%) in Experiment 2 and 3

			reading span	subtract-2	composite
Exp 2	Natives (N=19)	range	47.6 – 92.9	59.3 – 96.3	59.4 – 88.8
		mean	74.7	80.5	77.6
		SD	12.4	10.0	8.3
	L2 learners (N=22)	range	35.7 – 97.6	42.6 – 90.7	50.3 – 89.9
		mean	72.7	72.9	72.8
		SD	16.5	11.9	10.9
Exp 3	Natives (N=26)	range	42.9 - 100	61.1 - 100	62.4 – 100
		mean	79.2	87.2	83.2
		SD	15.8	9.7	10.5
	L2 learners (N=26)	range	19 – 97.6	55.6 - 87	44.7 – 86.8
		mean	56.8	72.2	64.5
		SD	18.4	8.8	11.5

In Experiment 3, the L2 learners' composite working memory scores were significantly lower than those of the native speakers (independent samples *t*-test: $p<.001$), unlike in Experiment 2, where the L2 learners' and the native speakers' composite working memory scores were not significantly different. This was due to the native speakers who participated in Experiment 3 had a significantly higher composite working memory scores than those who participated in Experiment 2, and the L2 learners who participated in Experiment 3 had a significantly lower composite working memory scores than those who participated in Experiment 2 (*ps*<.05). Therefore, comparisons of the effects related to working memory capacity across the two experiments must be made with caution.

6.1.2. Materials

Modeled on Phillips (2006), the study was designed to investigate the effects of two types of islands in *wh*-dependency formation. The islands were subject NPs with an infinitival complement (S-IC, hereafter) and subject NPs with a finite relative clause modifier (S-RC, hereafter). Four conditions were created by crossing the two types of subject NPs and two levels of Plausibility as illustrated below in (5).

(5) a. Subject NP with an infinitival complement (S-IC) / Plausible

Which schools will [the proposal to expand greatly on the curriculum] advance next year?

b. Subject NP with an infinitival complement (S-IC) / Implausible

Which students will [the proposal to expand greatly on the curriculum] overwhelm next year?

c. Subject NP with a finite relative clause modifier (S-RC) / Plausible

Which schools will [the proposal that expanded greatly on the curriculum] advance next year?

d. Subject NP with a finite relative clause modifier (S-RC) / Implausible

Which students will [the proposal that expanded greatly on the curriculum] overwhelm next year?

The sentences were *wh*-questions with a sentence-initial filler (*which schools/which students* in (5)). The filler must form a dependency with the main verb of the sentence (*advance/overwhelm*). However, another potential subcategorizer occurred earlier in the sentence (*expand*), which was embedded within the subject NP (The subject NP is marked with a square bracket in (5). The square bracket was not shown to the participants at the time of testing.). In order to investigate whether participants attempt to form a dependency at this first potential subcategorizer, plausibility was manipulated between the filler and this verb (it is plausible to expand schools but it is not plausible to expand students). Phillips' (2006) Experiment 1 showed that S-IC, but not S-RC, allows a parasitic gap. Note that the subject NP with a finite relative clause modifier is an island of the same type as the one whose effect was examined in Experiment 2.

The subject NPs were inanimate NPs that can take both an infinitival complement and a relative clause modifier (e.g., *plan*, *effort*, *campaign*, *proposal*). Within the islands, the first verbs after the filler (*expand*) were always followed by an adverb (*greatly*) and a PP (*on the curriculum*). As the first verbs, those that allow direct object complements were used (e.g., *expand the schools* is possible). This was to ensure that a dependency formed at the first verb does not violate the verb's subcategorization. The verbs, however, turned out to have a PP complement (*expand on the curriculum*). The earliest point in the sentence that unambiguously shows that the verb has a PP complement is somewhere in the middle of PP—the fact was not clear immediately at the preposition, since continuations such as *expand the schools on the north sides* are possible. All sentences were globally plausible at the main verb, which was the actual subcategorizer.

Twenty sets of experimental items were created, each with a true/false statement. The true/false statement did not directly ask where the *wh*-dependency should be formed. For example, the true/false statement for example (5) '*There was a proposal about the curriculum.*', for which the correct answer was 'True'. The answer for about half of the statements was true and the answer for the others was false. The twenty sets of items were distributed among four lists in a Latin square design, and each list also contained fifty-five fillers. The experimental and filler items were presented in a pseudorandom order. Appendix A shows the experimental items of Experiment 3.

6.1.3. Procedure

The participants completed the eye-tracking task first and proceeded to the other tasks in the order of the reading span test, the subtract-2 span test, the language background questionnaire and the cloze test. Experiment 3 was conducted under the same procedure as in Experiment 2.

6.1.4. Data analysis

The main purpose of Experiment 3 was to examine whether a gap is actively sought within the two types of islands. The critical regions were the first potential subcategorizer (*expand* in (5)), and the two following words as the spillover region (the adverb and the preposition (*greatly on*)). The preposition was included as the spillover region because the fact that the first verb has a PP complement was not yet clear at this point. In fact, it appears that the active gap search effect indeed spilled over to the preposition in Experiment 2. Moreover, it was deemed desirable to include as many words as possible as the spillover regions given that the experimental sentences were fairly complex.

As in Experiment 2, four eye movement measures (first-pass reading time, first-pass regression, regression path duration and total reading time) were computed and analyzed. Fixations were trimmed by the criteria used in Experiment 2. The data were statistically analyzed in linear/logistic mixed effects models with Plausibility (Plausible vs. Implausible), Island types (S-IC vs. S-RC), and Working Memory (the composite working memory scores) as the fixed effects. For more details, refer to the descriptions in Experiment 2.

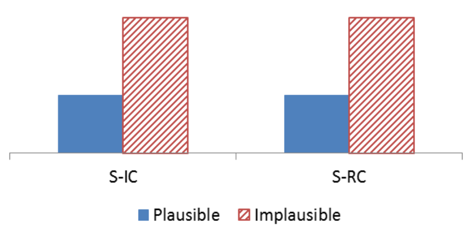
6.1.5. Prediction

The logic of the predictions is basically the same as before. The following predictions are based on the assumption that plausibility information is also considered when determining whether to form a wh-dependency (as well as grammatical information), as suggested by the results of Experiment 2.

1. If the grammar does not constrain online filler-gap dependency formation,

- Participants will attempt to form a dependency at the earliest potential subcategorizer (or at the spillover words) in S-IC and in S-RC. If the dependency is implausible, it will cause longer reading times/more first-pass regressions. Statistically, this will result in a significant main effect of Plausibility, without a significant interaction with Island type.

Figure 6.1. The pattern of eye movements predicted if the grammar does not constrain online wh-dependency formation

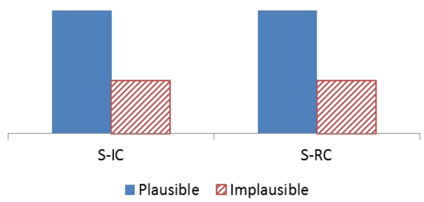


2. If island constraints constrain online filler-gap dependency formation, regardless of the possibility of parasitic gaps,

- Since both S-IC and S-RC are islands, grammatical information will not support the gap analysis in either island. However, the gap analysis is preferred in the Plausible condition.

Due to the competition between grammar and plausibility, the Plausible conditions will have longer reading times/more regressions than the Implausible conditions, regardless of island types. This will result in a significant main effect of Plausibility, without further interaction with Island type. Figure 6.2 shows the pattern.

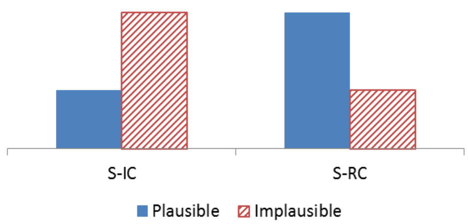
Figure 6.2. The pattern of eye movements predicted if island constraints constrain online filler-gap dependency formation, but the grammar of parasitic gaps does not



3. If both island constraints and the possibility of allowing parasitic gaps influence online filler-gap dependency formation

- In S-IC, forming a wh-dependency is a grammatically allowed option. This information will compete with the plausibility information in the Implausible condition, causing longer reading times/more regressions. In S-RC, which does not allow a gap, the Plausible condition will have longer reading times/more regressions. A pattern similar to Figure 6.3 will therefore be found. Statistically, there will be a significant interaction between Plausibility and Island type.

Figure 6.3. The pattern of eye movements predicted if both island constraints and the possibility of allowing parasitic gaps influence online filler-gap dependency formation



6.2. Results

Comprehension question accuracy

Mean comprehension question accuracy of the native speaker group was 90.8% (range=75.3%-100%, SD=6.9%), suggesting that they paid attention to the task. Examination of the L2 learners' data revealed that one participant's mean accuracy score was close to chance level (56.8%), and this participant's data was excluded from the main analyses. The mean comprehension question accuracy of the remaining 26 L2 learners was 84% (range=65%-95.2%, SD= 8.7%).

6.2.1. Native speakers' eye movement

Table 6.3 shows a descriptive summary of the native speakers' eye movement at the critical words. The data from each word were analyzed in mixed effects models. The fixed effects of the models are presented in Tables 6.4 – 6.6.

Table 6.3. Means and standard deviations of the eye movement measures at the critical words: Native speakers

	first possible	adverb	preposition
First-pass reading time			
S-IC/plausible	300 (156)	323 (159)	265 (102)
S-IC/implausible	293 (135)	323 (155)	281 (128)
S-RC/plausible	273 (154)	269 (107)	249 (96)
S-RC/implausible	273 (144)	297 (140)	250 (99)
First-pass regression			
S-IC/plausible	0.33 (0.47)	0.15 (0.36)	0.16 (0.37)
S-IC/implausible	0.38 (0.49)	0.17 (0.37)	0.16 (0.37)
S-RC/plausible	0.38 (0.49)	0.24 (0.43)	0.07 (0.26)
S-RC/implausible	0.23 (0.42)	0.25 (0.43)	0.08 (0.28)
Regression path duration			
S-IC/plausible	558 (531)	454 (465)	393 (548)
S-IC/implausible	622 (863)	411 (248)	396 (362)
S-RC/plausible	582 (708)	500 (738)	287 (225)
S-RC/implausible	413 (367)	504 (628)	359 (500)
Total reading time			
S-IC/plausible	627 (411)	654 (454)	271 (280)
S-IC/implausible	654 (490)	672 (411)	324 (331)
S-RC/plausible	700 (526)	578 (337)	257 (261)
S-RC/implausible	660 (520)	607 (452)	274 (267)

Table 6.4. Fixed effects of the model fitted to the measures of native speakers' eye movement at the first possible subcategorizer (*expand*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	282.771	13.712	20.622	0.000	***
	Plausibility	-3.631	12.513	-0.29	0.778	
	Island	-21.974	12.529	-1.754	0.075	†
	WM	1.089	1.214	0.897	0.325	
	Plausibility:Island	10.227	25.065	0.408	0.681	
	Plausibility:WM	0.931	1.212	0.768	0.435	
	Island:WM	-0.564	1.239	-0.456	0.611	
	Plausibility:Island:WM	-0.078	2.481	-0.031	0.974	
First-pass regression	(Intercept)	-0.915	0.259	-3.53	0.000	***
	Plausibility	-0.313	0.222	-1.413	0.158	
	Island	-0.240	0.222	-1.086	0.278	
	WM	-0.01	0.025	-0.465	0.642	
	Plausibility:Island	-1.113	0.444	-2.506	0.012	*
	Plausibility:WM	-0.014	0.022	-0.624	0.533	
	Island:WM	0.053	0.022	2.388	0.017	*
	Plausibility:Island:WM	-0.009	0.044	-0.207	0.836	
Regression path duration	(Intercept)	541.772	63.422	8.542	0.000	***
	Plausibility	-57.246	52.981	-1.081	0.281	
	Island	-70.196	53.018	-1.324	0.186	
	WM	-4.36	6.067	-0.719	0.473	
	Plausibility:Island	-236.568	106.028	-2.231	0.026	*
	Plausibility:WM	-2.82	5.13	-0.55	0.583	
	Island:WM	9.606	5.162	1.861	0.063	†
	Plausibility:Island:WM	-4.113	10.323	-0.398	0.691	
Total reading time	(Intercept)	660.708	56.664	11.66	0.000	***
	Plausibility	-7.248	36.544	-0.198	0.843	
	Island	40.448	36.587	1.106	0.270	
	WM	-4.003	4.678	-0.856	0.393	
	Plausibility:Island	-68.523	73.18	-0.936	0.350	
	Plausibility:WM	3.24	3.533	0.917	0.360	
	Island:WM	1.849	3.677	0.503	0.615	
	Plausibility:Island:WM	-3.211	7.359	-0.436	0.663	

At the first possible subcategorizer (*expand*), there was a marginally significant main effect of Island in the first-pass reading time, reflecting that the first-pass reading time was longer in the S-IC than in the S-RC. The effect of Island was modulated by Working Memory in the first-pass regression and there was a similar trend in the regression path duration. This was because the lower span native speakers made more first-pass regressions in and took longer to read S-IC ([first-pass regression] S-IC: 0.35, S-RC: 0.22, [regression path duration] S-IC: 625ms, S-RC: 432ms), whereas the higher span speakers did not ([first-pass regression] S-IC: 0.36, S-RC: 0.39, [regression path duration] S-IC: 552ms, S-RC: 561ms).

There was also a significant interaction between Plausibility and Island in the first-pass regression and regression path duration, due to the different directions of the plausibility effect in the two types of island. As shown in Figure 6.4 below, in the S-IC condition it was the Implausible condition that induced a higher proportion of first-pass regression/longer regression path duration whereas in the S-RC condition it was the Plausible condition. For both the first-pass regression and regression path duration, the plausibility effect was bigger in S-RC, due to a noticeably lower proportion of first-pass regressions/shorter regression path duration of the S-RC/Implausible condition.

Figure 6.4. (a) first-pass regression and (b) regression path duration in the four conditions at the first subcategorizer

(a) first-pass regression (b) regression path duration

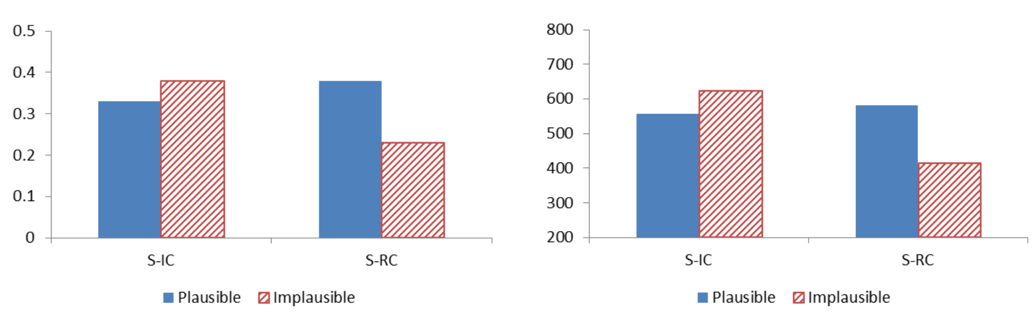


Table 6.5. Fixed effects of the model fitted to the measures of native speakers' eye movement at the adverb immediately following the first possible subcategorizer (*greatly*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	302.032	13.053	23.139	0.0001	***
	Plausibility	13.750	11.713	1.174	0.252	
	Island	-39.078	11.72	-3.334	0.001	**
	WM	1.355	0.888	1.526	0.120	
	Plausibility:Island	27.885	23.442	1.19	0.236	
	Plausibility:WM	0.793	1.129	0.702	0.489	
	Island:WM	0.995	1.175	0.847	0.411	
	Plausibility:Island:WM	0.495	2.346	0.211	0.810	
First-pass regression	(Intercept)	-1.459	0.142	-10.284	0.000	***
	Plausibility	0.067	0.231	0.288	0.773	
	Island	0.540	0.231	2.339	0.019	*
	WM	-0.013	0.014	-0.948	0.343	
	Plausibility:Island	-0.105	0.462	-0.228	0.820	
	Plausibility:WM	-0.015	0.022	-0.68	0.497	
	Island:WM	-0.019	0.022	-0.857	0.392	
	Plausibility:Island:WM	-0.008	0.044	-0.187	0.851	
Regression path duration	(Intercept)	467.544	49.570	9.432	0.000	***
	Plausibility	-24.747	45.163	-0.548	0.584	
	Island	72.960	45.144	1.616	0.107	
	WM	-2.544	4.790	-0.531	0.596	
	Plausibility:Island	39.788	90.310	0.441	0.660	
	Plausibility:WM	-0.098	4.353	-0.023	0.982	
	Island:WM	-9.195	4.350	-2.114	0.035	*
	Plausibility:Island:WM	9.089	8.701	1.045	0.297	
Total reading time	(Intercept)	627.943	43.881	14.31	0.000	***
	Plausibility	23.184	32.454	0.714	0.475	
	Island	-68.400	32.489	-2.105	0.036	*
	WM	0.176	3.659	0.048	0.962	
	Plausibility:Island	12.435	64.983	0.191	0.848	
	Plausibility:WM	1.286	3.138	0.41	0.682	
	Island:WM	-0.224	3.252	-0.069	0.945	
	Plausibility:Island:WM	1.018	6.508	0.156	0.876	

At the adverb immediately following the first subcategorizer, there was a significant main effect of Island in the first-pass reading time due a longer reading time in S-IC than S-RC (S-IC: 323ms, S-RC: 283ms). There was also a significant main effect of Island in the first-pass regression but in the reverse direction (S-IC: 0.16, S-RC: 0.24). The main effect of Island interacted with Working Memory in the regression path duration and this was mainly because the lower span group read S-RC longer than S-IC ([higher span] S-IC: 448ms, S-RC: 459ms, [lower span] S-IC: 417ms, S-RC: 547ms). In the total reading time, there was a significant main effect of Island, due to a longer reading time in S-IC (S-IC: 663ms, S-RC: 593ms)

Table 6.6. Fixed effects of the model fitted to the measures of native speakers' eye movement at the preposition (*on*)

		Estimate	Std. error	*t*/*z* value	*p* value	
First-pass reading time	(Intercept)	263.490	11.367	23.18	0.0001	***
	Plausibility	8.614	13.448	0.64	0.547	
	Island	-25.914	13.566	-1.91	0.051	†
	WM	1.071	0.661	1.62	0.140	
	Plausibility:Island	-8.915	27.026	-0.33	0.782	
	Plausibility:WM	-0.561	1.329	-0.422	0.656	
	Island:WM	-1.068	1.364	-0.783	0.399	
	Plausibility:Island:WM	4.302	2.702	1.592	0.107	
First-pass regression	(Intercept)	-2.303	0.286	-8.042	0.000	***
	Plausibility	0.009	0.481	0.019	0.985	
	Island	-0.908	0.481	-1.889	0.059	†
	WM	-0.020	0.027	-0.721	0.471	
	Plausibility:Island	-0.002	0.961	-0.002	0.999	
	Plausibility:WM	-0.080	0.047	-1.709	0.088	†
	Island:WM	-0.039	0.046	-0.857	0.391	
	Plausibility:Island:WM	0.026	0.092	0.287	0.774	
Regression path duration	(Intercept)	359.52	36.563	9.833	0.000	***
	Plausibility	48.707	54.63	0.892	0.374	
	Island	-81.118	54.572	-1.486	0.139	
	WM	-3.858	3.567	-1.081	0.281	
	Plausibility:Island	49.936	109.149	0.458	0.648	
	Plausibility:WM	-7.246	5.362	-1.351	0.178	
	Island:WM	1.731	5.329	0.325	0.746	
	Plausibility:Island:WM	1.734	10.672	0.162	0.871	
Total reading time	(Intercept)	281.504	32.077	8.776	0.000	***
	Plausibility	35.061	21.837	1.606	0.109	
	Island	-33.167	21.865	-1.517	0.130	
	WM	1.617	2.487	0.65	0.516	
	Plausibility:Island	-39.857	43.733	-0.911	0.363	
	Plausibility:WM	2.691	2.111	1.274	0.203	
	Island:WM	-1.333	2.201	-0.605	0.545	
	Plausibility:Island:WM	1.707	4.406	0.387	0.699	

At the preposition, there was a marginally significant main effect of Island in the first-pass reading time and in the first-pass regression, due to a longer first-pass reading time in S-IC than in S-RC (S-IC: 273ms, S-RC: 249ms) and more first-pass regressions in S-IC than in S-RC (S-IC: 0.16, S-RC: 0.08). There was also a trend toward interaction between Plausibility and Working Memory, reflecting that the high span native speakers made more regressions in the Plausible than in the Implausible conditions (Plausible: 0.16, Implausible: 0.09) and the lower span native speakers showed a reverse trend (Plausible: 0.07, Implausible: 0.16).

6.2.2. L2 learners' eye movements

A summary of L2 learners' eye movements is in Table 6.7. The fixed effects of the mixed effects models fitted to the data from each critical word are presented in Tables 6.8 to 6.10.

Table 6.7. Means and standard deviations of the eye movement measures at the critical words: L2 learners

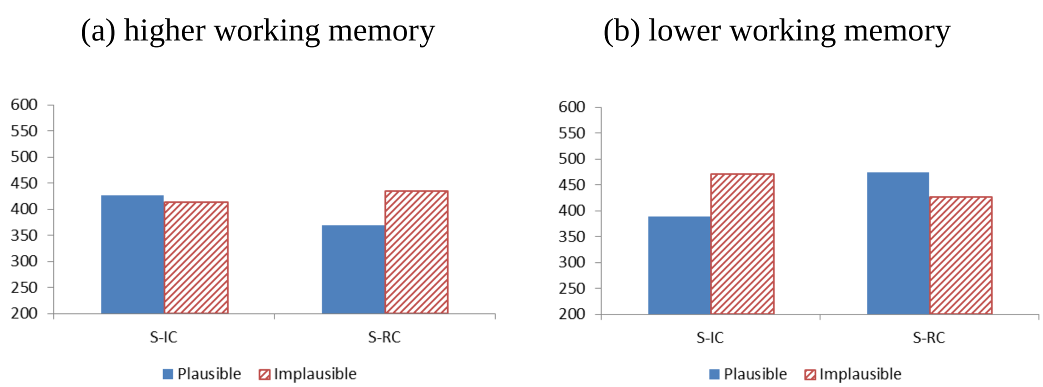
	first possible	adverb	preposition
First-pass reading time			
S-IC/plausible	401 (220)	459 (292)	294 (192)
S-IC/implausible	444 (247)	418 (235)	319 (158)
S-RC/plausible	431 (260)	436 (236)	291 (111)
S-RC/implausible	429 (237)	433 (248)	305 (109)
First-pass regression			
S-IC/plausible	0.33 (0.47)	0.24 (0.43)	0.10 (0.31)
S-IC/implausible	0.25 (0.43)	0.23 (0.42)	0.19 (0.39)
S-RC/plausible	0.30 (0.46)	0.20 (0.40)	0.19 (0.40)
S-RC/implausible	0.30 (0.46)	0.19 (0.40)	0.09 (0.29)
Regression path duration			
S-IC/plausible	788 (709)	864 (1042)	529 (940)
S-IC/implausible	922 (1552)	801 (919)	511 (520)
S-RC/plausible	804 (733)	780 (1030)	473 (775)
S-RC/implausible	988 (1823)	696 (709)	385 (296)
Total reading time			
S-IC/plausible	1129 (822)	1040 (693)	428 (436)
S-IC/implausible	1118 (778)	1082 (704)	435 (450)
S-RC/plausible	1207 (867)	1020 (771)	414 (380)
S-RC/implausible	1275 (846)	1001 (590)	458 (428)

Table 6.8. Fixed effects of the model fitted to the measures of L2 learners' eye movement at the first possible subcategorizer (*expand*)

		Estimate	Std. error	*t*/*z* value	*p* value	
First-pass reading time	(Intercept)	422.985	26.907	15.72	0.0001	
	Plausibility	20.880	19.255	1.084	0.274	
	Island	8.951	19.272	0.464	0.661	
	WM	-1.665	1.557	-1.07	0.254	
	Plausibility:Island	-54.016	38.442	-1.405	0.163	
	Plausibility:WM	-0.176	2.024	-0.087	0.963	
	Island:WM	-2.481	1.745	-1.421	0.166	
	Plausibility:Island:WM	9.123	3.477	2.624	0.010	*
First-pass regression	(Intercept)	-0.937	0.130	-7.236	0.000	***
	Plausibility	-0.210	0.205	-1.028	0.304	
	Island	0.081	0.205	0.395	0.693	
	WM	0.006	0.012	0.474	0.636	
	Plausibility:Island	0.489	0.409	1.194	0.232	
	Plausibility:WM	0.012	0.019	0.624	0.533	
	Island:WM	-0.030	0.019	-1.594	0.111	
	Plausibility:Island:WM	-0.056	0.037	-1.485	0.138	
Regression path duration	(Intercept)	866.271	100.819	8.592	0.000	***
	Plausibility	160.720	113.006	1.422	0.156	
	Island	44.261	113.076	0.391	0.696	
	WM	-4.097	8.365	-0.49	0.625	
	Plausibility:Island	51.459	225.874	0.228	0.820	
	Plausibility:WM	4.647	10.748	0.432	0.666	
	Island:WM	0.008	10.233	0.001	0.999	
	Plausibility:Island:WM	27.372	20.417	1.341	0.181	
Total reading time	(Intercept)	1178.266	96.772	12.176	0.000	***
	Plausibility	38.033	59.047	0.644	0.520	
	Island	110.165	58.934	1.869	0.062	†
	WM	-19.85	7.261	-2.734	0.007	**
	Plausibility:Island	85.108	117.885	0.722	0.471	
	Plausibility:WM	-9.047	6.06	-1.493	0.136	
	Island:WM	-8.239	5.277	-1.561	0.119	
	Plausibility:Island:WM	12.511	10.549	1.186	0.236	

At the first subcategorizer, the model fitted to the first-pass reading times revealed a significant interaction between Plausibility, Island and Working Memory. The source of the three-way interaction was explored by examining the first-pass reading time of each working memory group separately as shown in Figure 6.5.

Figure 6.5. First-pass reading times of (a) the higher span L2 learners and (b) the lower span L2 learners at the first subcategorizer



The other significant fixed effects at this region was a marginally significant main effect of Island in total reading time, due to a longer total reading time of S-RC (S-IC: 1123ms, S-RC: 1241ms), and a significant main effect of Working Memory, reflecting that the lower span learners' total reading time was longer than the higher span learners' total reading time (higher span learners: 964ms, lower span learners: 1363ms)

Table 6.9. Fixed effects of the model fitted to the measures of L2 learners' eye movement at the adverb immediately following the first possible subcategorizer (*greatly*)

		Estimate	Std. error	*t*/*z* value	*p* value	
First-pass reading time	(Intercept)	434.388	27.154	15.997	0.0001	***
	Plausibility	-20.121	20.401	-0.986	0.328	
	Island	-7.074	20.378	-0.347	0.739	
	WM	-2.176	1.726	-1.26	0.194	
	Plausibility:Island	34.403	40.697	0.845	0.397	
	Plausibility:WM	0.960	2.120	0.453	0.645	
	Island:WM	-3.103	1.846	-1.681	0.099	†
	Plausibility:Island:WM	1.400	3.689	0.379	0.705	
First-pass regression	(Intercept)	-1.385	0.163	-8.505	0.000	***
	Plausibility	-0.093	0.227	-0.408	0.683	
	Island	-0.254	0.228	-1.114	0.265	
	WM	-0.011	0.013	-0.854	0.393	
	Plausibility:Island	0.014	0.455	0.032	0.975	
	Plausibility:WM	-0.020	0.022	-0.876	0.381	
	Island:WM	-0.034	0.021	-1.587	0.113	
	Plausibility:Island:WM	0.012	0.042	0.294	0.768	
Regression path duration	(Intercept)	786.008	73.689	10.667	0.000	***
	Plausibility	-70.222	80.332	-0.874	0.383	
	Island	-97.349	80.273	-1.213	0.226	
	WM	-5.108	5.809	-0.879	0.380	
	Plausibility:Island	-5.853	160.411	-0.036	0.971	
	Plausibility:WM	1.436	7.813	0.184	0.854	
	Island:WM	-5.354	7.269	-0.737	0.462	
	Plausibility:Island:WM	-6.435	14.53	-0.443	0.658	
Total reading time	(Intercept)	1032.328	75.961	13.59	0.000	***
	Plausibility	12.039	51.353	0.234	0.815	
	Island	-58.197	51.258	-1.135	0.257	
	WM	-15.585	5.579	-2.793	0.005	**
	Plausibility:Island	-55.242	102.531	-0.539	0.590	
	Plausibility:WM	-3.591	5.249	-0.684	0.494	
	Island:WM	-3.692	4.589	-0.804	0.422	
	Plausibility:Island:WM	5.826	9.175	0.635	0.526	

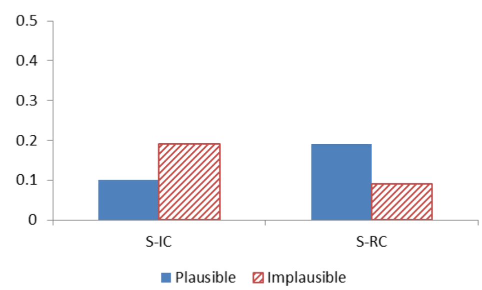
At the adverb immediately following the first subcategorizer, there was a marginally significant interaction between Island and Working Memory in the first-pass reading time. This was because the higher span learners read S-IC slower than S-RC in the first pass (S-IC: 422ms, S-RC: 391ms) whereas the lower span learners showed the opposite tendency (S-IC: 452ms, S-RC: 472ms). In the total reading time, the main effect of Working Memory was significant, due to a longer total reading time of the lower span learners than the higher span learners (higher span: 839ms, lower span: 1199ms).

Table 6.10. Fixed effects of the model fitted to the measures of L2 learners' eye movement at the preposition (*on*)

		Estimate	Std. error	*t/z* value	*p* value	
First-pass reading time	(Intercept)	284.070	17.918	15.854	0.0001	***
	Plausibility	9.057	15.124	0.599	0.513	
	Island	-6.136	15.21	-0.403	0.691	
	WM	-2.104	1.213	-1.735	0.065	†
	Plausibility:Island	1.976	30.513	0.065	0.999	
	Plausibility:WM	0.140	1.595	0.088	0.989	
	Island:WM	0.687	1.379	0.498	0.635	
	Plausibility:Island:WM	3.244	2.733	1.187	0.268	
First-pass regression	(Intercept)	-2.135	0.289	-7.38	0.000	***
	Plausibility	-0.060	0.392	-0.153	0.879	
	Island	-0.045	0.393	-0.114	0.909	
	WM	-0.043	0.021	-2.052	0.040	*
	Plausibility:Island	-1.911	0.790	-2.419	0.016	*
	Plausibility:WM	0.018	0.042	0.436	0.663	
	Island:WM	-0.009	0.038	-0.229	0.819	
	Plausibility:Island:WM	-0.063	0.077	-0.825	0.409	
Regression path duration	(Intercept)	469.43	53.418	8.788	0.000	***
	Plausibility	-61.96	77.441	-0.8	0.424	
	Island	-92.797	77.693	-1.194	0.233	
	WM	-7.956	3.491	-2.279	0.023	*
	Plausibility:Island	-73.247	155.96	-0.47	0.639	
	Plausibility:WM	1.683	7.685	0.219	0.827	
	Island:WM	1.681	7.037	0.239	0.811	
	Plausibility:Island:WM	7.383	14.049	0.526	0.600	
Total reading time	(Intercept)	432.335	57.254	7.551	0.000	***
	Plausibility	21.689	29.844	0.727	0.468	
	Island	7.849	29.769	0.264	0.792	
	WM	-5.035	3.085	-1.632	0.103	
	Plausibility:Island	44.374	59.545	0.745	0.457	
	Plausibility:WM	-3.708	3.163	-1.173	0.242	
	Island:WM	3.095	2.666	1.161	0.246	
	Plausibility:Island:WM	5.641	5.329	1.059	0.290	

At the preposition, the main effect of Working Memory was marginally significant in the first-pass reading time, and significant in the first-pass regression and regression path duration. All these effects indicate that the lower span learners took more time to read this region and made more first-pass regressions from this region than the higher span learners. There was also a significant interaction between Plausibility and Island in first-pass regressions. As Figure 6.6 shows, this was because the Implausible condition induced a higher proportion of first-pass regression than the Plausible condition in S-IC, whereas it was vice versa in the S-RC.

Figure 6.6. L2 learners' proportions of the first-pass regression in the four conditions at the preposition



6.3. Discussion of the native speakers' and L2 learners' eye movements

The native speaker data showed a significant interaction of Plausibility and Island type in first-pass regression and regress-path duration, replicating the finding of Phillips' (2006) self-paced reading study with a different methodology. The results suggest that native speakers' active gap search works differently not only between the sentences with and without an island (Experiment 2) but also between the two kinds of subject islands (Experiment 3). This finding would constitute evidence that native speakers' real-time sentence processing is constrained by knowledge about which islands allow a parasitic gap and which islands do not.

As for the L2 learners, two regions revealed the effects that speak to the learners' sensitivity to the grammar: at the first verb and at the preposition two words downstream. At the first verb region, there was a significant three-way interaction of Plausibility, Island and Working Memory in first-pass reading time; at the preposition, the interaction of Plausibility and Island reached significance in first-pass regression.

Figure 6.5 illustrates the pattern of the three-way interaction of Plausibility, Island and Working Memory in the L2 learners' first-pass reading time. As the figure shows, it was the lower span learners that showed evidence that the types of island and the corresponding parasitic gap distribution are taken into account in the early stages of processing. In Experiment 2, the higher span native speakers tended to apply the island constraints earlier than the lower span counterparts, and the L2 learners showed a numerically similar pattern. Based on this finding, it was hypothesized that higher L2 working memory capacity is associated with faster application of grammatical information in L2 sentence processing. This hypothesis, however, is not consistent with the result of Experiment 3.

The apparently conflicting results of Experiment 2 and 3 might be accounted for if the lower working memory learners have not yet acquired the grammatical knowledge with regard to the different types of islands and the corresponding parasitic gap distribution. The sensitivity of the lower working memory learners to the island types at the first verb might then be attributed to the difference in the degree of processing difficulty between the embedded subject NPs in the S-IC conditions (e.g., *the proposal to expand greatly on curriculum*) and those in the S-RC conditions (e.g., *the proposal that expanded greatly on the curriculum*). According to the view that finiteness of the verb increases processing loads (e.g., Kluender, 2004), the complex NP that contains a finite relative clause would be harder to process than the one that contains an infinitival complement. Along the line of the processing accounts of islands, therefore, it will be harder to carry out gap search within an S-RC than in an S-IC. The apparent sensitivity of the lower working memory learners to the island types could then be attributed to their sensitivity to the relative processing difficulty of the two types of island constructions, which is ultimately due to their lack of relevant grammatical knowledge.

Note that this processing-based explanation does not apply to the learners with higher working memory capacity because a strong version of processing-based accounts of island (e.g., Hofmeister & Sag, 2010; Kluender, 1998, 2004; Kluender & Kutas, 1993) would predict that individuals with higher working memory will be more likely to evaluate the possibility of gap postulation within both S-IC and S-RC than those with lower working memory capacity, but in this experiment they did not (see Figure 6.5a). Given that there was a significant interaction of Plausibility and Island in the first-pass regression at the preposition, the results of Experiment 3 are more consistent with the possibility that at least some L2 learners (probably those with higher working memory capacity) have acquired the native-like grammatical knowledge about island constraints and parasitic gap distribution and can apply this knowledge in online filler-gap processing though with slower speed as compared to native speakers. Admittedly, this interpretation is speculative at the moment and further research is in order into the way L2 learners utilize constraints on parasitic gaps in filler-gap processing and the role of working memory capacity therein.

Also noteworthy is that the overall pattern of the interaction of Plausibility and Island in Experiment 3 looks more similar to what was shown in Experiment 2 than in Experiment 1. The result of Experiment 1 was interpreted to be consistent with the possibility that the island constraints work deterministically in preventing filler-gap dependency formation regardless of its semantic plausibility. The results of Experiments 2 and 3, in contrast, are more consistent with the pattern that would be expected if the grammatical constraint and semantic plausibility simultaneously exert an influence on the implementation of active gap search. Taken together, the results of the three experiments lend support to the hypothesis that the nature of the effect that different types of information have on active gap search processes is affected by the types of experimental tasks (stop-making-sense task vs. eye-tracking). The present results suggest that in unmonitored and relatively faster reading, which is more likely to be induced by eye-tracking than by the stop-making-sense task, the effect of grammatical constraints may not always override the influences from semantic constraints.

Chapter 7

General discussion and conclusion

The present dissertation aimed to advance our understanding of the characteristics of the sentence processing mechanisms that adult L2 learners employ to comprehend L2 sentences, by investigating their online processing of long-distance syntactic dependencies between a displaced wh-phrase (filler) and its original licensor (gap) as well as their sensitivity to grammatical constraints on the formation of such dependencies. The focus was on the sentences that contained syntactic relations between a filler and its gap. Between the filler and the gap was a potential, but ultimately incorrect gap site that was either within a syntactic island or not. An online plausibility judgment study and two eye tracking studies were conducted to investigate how native English speakers and Korean L2 learners of English process filler-gap dependencies and whether they are sensitive to island constraints. In this last chapter, the major findings of the present experiments are first presented and then they are discussed in terms of major issues in L2 sentence processing and grammar acquisition.

In Experiment 1, a stop-making-sense task was conducted to investigate L2 learners' sensitivity to the subject/relative clause island constraint in online sentence processing. The native speakers showed evidence of immediately applying the island constraint. That is, although they interpreted a wh-dependency at the earliest possible gap site when it is grammatically licit, they suspended the immediate gap postulation when they detected a syntactic island. The reading time data showed that the L2 learners were not so successful in immediately suppressing the urge to complete a wh-dependency. However, in the plausibility judgment data the learners showed evidence of being capable of ultimately ruling out an illegal dependency in accordance with the island constraint. Experiment 1 also showed that when an initially formed wh-dependency turns out to be incorrect, the native speakers attempted to reanalyze the initial parse, but the L2 learners failed to do so.

Experiment 2 employed eye-movement monitoring to examine the way the native speakers and L2 learners apply the subject/relative clause island constraint when processing filler-gap dependencies under a more natural reading situation. Working memory capacity of the participants was also collected in an attempt to capture potential individual differences in filler-gap processing and grammar application. The results indicate that even native speakers sometimes fail to completely suppress automatic active gap creation inside an island at least in early stages of processing. However, they eventually showed evidence of successful application of the island constraint. The learners' eye-movement patterns also showed that they were able to apply island constraints in the last-pass reading. There was also suggestive evidence that readers with larger working memory capacity apply island constraints earlier than those with smaller working memory capacity. As in Experiment 1, the native speakers showed evidence of attempting to reanalyze the incorrect initial wh-dependency, but the L2 learners did not show such evidence. The experiment also revealed speed differences in constructing a relative clause between the native speakers and the L2 learners, and among learners with different working memory capacity.

Lastly, Experiment 3 investigated sensitivity to the differential distribution of parasitic gaps within two kinds of syntactic islands, one of which allows a parasitic gap and the other does not. The native speakers rapidly distinguished two types of islands, and performed active gap search in only the island that allows a parasitic gap. Some of the L2 learners showed a similar pattern of sensitivity to this subtle grammatical constraint, though it appeared in a later region than the native speakers.

The results of the present experiments summarized above offer several implications for understanding the architecture of the L2 sentence processing mechanism in comparison to the L1 sentence processing mechanism. Issues that have been addressed in previous research on the nature of L2 sentence processing include the characteristics of L2 syntactic representations underlying L2 sentence processing (e.g., Clahsen & Felser, 2006a, b, c; Hopp, 2010), the role of working memory capacity in L2 sentence comprehension (e.g., Dussias & Piñar, 2010; Felser & Roberts, 2007; Havik et al., 2009), and the interaction between syntactic and non-syntactic information in the course of syntactic reanalysis (e.g., Dussias & Piñar, 2010; Roberts & Felser, 2011; Williams et al., 2001). This dissertation provides empirical findings relevant to these issues, by testing participants with different working memory capacity on the online applicability of island constraints using plausibility effects as a diagnostic. We first discuss what the present results suggest about the characteristics of grammatical representations computed in L2 sentence processing.

As mentioned before, a growing number of experimental findings suggest that adult L2 learners may have a problem in using morpho-syntactic information in real time L2 sentence processing compared to lexical, semantic or pragmatic information (e.g., Dussias & Cramer Scaltz, 2008; Felser et al., 2003, 2012; Felser & Roberts, 2007; Jiang, 2004; Marinis et al., 2005; Weber-fox & Neville, 1996, 2001). These findings have provided an empirical basis for the Shallow Structure Hypothesis (Clahsen & Felser, 2006a, b, c), according to which adult learners' L2 sentence processing is characterized by the reduced ability to conduct full syntactic analysis. The L2 parser thus ends up relying on shallow processing which primarily resorts to non-syntactic information to comprehend sentences. Under the assumption that native speakers typically compute full syntactic analysis under normal sentence processing circumstances, the hypothesis emphasizes fundamental differences in the nature of L1 and L2 processing. In the sense that the hypothesis advocates a categorical distinction between L1 and L2 grammatical representations, it could be seen as a processing version of the view that the grammatical knowledge learned by late L2 learners has representational deficits (e.g., Bley-Vroman, 1990). According to the Shallow Structure Hypothesis, the differences between L1 and L2 sentence processing are *qualitative*, which cannot be explained in terms of *quantitative* differences between L1 and L2 speakers such as their working memory capacity, proficiency in the target language or speed of processing.

Examining how L2 learners apply island constraints in real time processing has a bearing on the questions about the nature of L2 grammatical processing. The three experiments conducted in this dissertation consistently yielded evidence that L2 learners, like native speakers, are sensitive to island constraints at least at some point during the course of online filler-gap processing, although island effects often appeared either in later regions or in late measures for the L2 learners compared to the native speakers. In Experiment 1, island constraints ultimately ruled out an ungrammatical wh-dependency when L2 learners performed online plausibility judgments. In Experiment 2, L2 learners' eye movement patterns showed online sensitivity to island constraints in last-pass reading even when they read sentences naturally without monitoring semantic plausibility. Experiment 3 showed that (at least some) L2 learners were also sensitive to the distinction between the two kinds of islands that differ in their potential to license an internal parasitic gap, providing suggestive evidence that L2 learners are able to apply highly sophisticated grammatical knowledge to sentence processing. Overall, proficient late learners' processing profiles were similar to native speakers' processing profiles in the sense that they were able to apply grammatical constraints to guide online parsing decisions as to whether to form a wh-dependency or not. All three experiments therefore lend support to the possibility that late L2 learners can in principle learn the target language's grammatical constraints to the extent that they can deploy them during real time comprehension. These findings cast doubts on a categorical distinction between L1 and L2 grammatical knowledge in terms of its availability to real-time sentence processing, undermining the view that late-learned L2 grammatical representations and their use in online processing are qualitatively different from those in the native language.

Examining the online application of island constraints is a good way to test syntactic precision in L2 processing. According to the structural description of island constraints, the island effects occur because a licit, long-distance syntactic association between a filler and a gap in a sentence cannot cross the boundary of a certain domain that is defined by the syntactic configuration of the words within the domain (e.g., Chomsky, 1986; Lasnik & Saito, 1992). Therefore, obeying island constraints in real-time comprehension would imply a successful computation of the correct syntactic configuration, including hierarchical structures, as well as a successful application of the constraint over the configuration. L2 learners' online sensitivity to the islands in this dissertation would thus provide evidence that the L2 learners were able to construct structural representations detailed enough to project islands, which again is not compatible with the Shallow Structure Hypothesis. Note, however, that the adequacy of the inferred link between L2 learners' sensitivity to the island and the depth of their syntactic analysis depends on the assumption that the island phenomena result from the grammatical constraints operating on structurally defined domains. Whether the source of the island phenomena lies in the grammar or the capacity-limited processing system is an ongoing debate (e.g., Hofmeister & Sag, 2010; Kluender, 1998, 2004; Phillips, 2006; Spouse, Wagers, & Phillips, 2012), and the issue will be further discussed later in this section.

Another piece of evidence that points to the qualitative similarity between L1 and L2 sentence processing comes from the finding of Experiment 2 that even native speakers may rely on ‘shallow’ application of island constraints under certain circumstances. In Experiment 2, there was a significant main effect of Plausibility two words after the first verb in the first-pass regression and regression path duration. This is suggestive evidence that native speakers temporarily considered an illegal gap analysis within the island, although they applied island constraints at a later stage (as shown by the interaction of Plausibility and Island at the first verb and at the following word in the total reading time). The results suggest that even native speakers did not deploy island constraints in an early stage of processing. Also notable is that although the stop-making-sense task in Experiment 1 and the eye-movement monitoring in Experiment 2 used very similar materials, native speakers showed evidence for shallow processing only in the latter. This observation suggests that in normal situations of reading for comprehending meaning, native speakers also may resort to shallow processing especially when coping with complex structures. This finding is in line with the research that showed incomplete syntactic analysis and temporary entertainment of ungrammatical representations in native language processing (e.g., Christianson et al., 2001; Ferreira, 2003; Tabor, Galantucci & Richardson, 2004). It seems that shallow processing strategies are not something peculiar to L2 sentence processing, but should be seen as a universal aspect of human sentence processing.

Another area in which qualitative similarity between L1 and L2 processing can be inferred is in the effect of individual differences in working memory capacity in online application of island constraints. In Experiment 2, there was a three-way interaction between Plausibility, Island and Working Memory in regression path duration at the adverb following the first possible subcategorizer in the native speaker group. This three-way interaction occurred because the higher span native speakers showed sensitivity to the island but the lower span native speakers posited an ungrammatical gap within the island. The results suggest that native speakers' sensitivity to island constraints may be modulated by their working memory capacity. What is notable here is that the L2 learners showed a similar trend at the same region in the same eye-movement measure, though it did not reach statistical significance. Although the lack of statistical robustness does not allow us to make a strong conclusion at this point, the results suggest an interesting possibility that individual differences in working memory capacity may exert a similar influence on L1 and L2 sentence processing in terms of how efficiently island constraints are applied. Future research needs to test more participants to see whether the working memory capacity effect is a reliable phenomenon in L2 processing.

Summing up, adult L2 learners' online processing of filler-gap dependencies and online application of island constraints display multiple qualitative similarities to L1 processing. First, the L2 learners were able to ultimately apply island constraints during online processing of filler-gap dependencies, and it can be inferred from the sensitivity to island constraints that they can in principle compute syntactic representations detailed enough to project island configurations. Second, shallow application of island constraints is not a characteristic property of L2 processing alone, and native speakers' initial stages of processing also seem to resort to shallow processing under certain circumstances. Finally, there was a suggestive trend that both L1 and L2 speakers' sensitivity to the island is modulated by individual differences in working memory capacity. These observations do not seem to fit well with the underlying idea of the Shallow Structure

Hypothesis that there is a qualitative difference between L1 and L2 grammatical representations, that adult L2 learners do not construct detailed enough syntactic parse, and that quantitative variables such as working memory capacity cannot explain the differences between L1 and L2 processing. Given the findings of this dissertation and other previous works that showed native-like sentence processing performance of adult L2 learners (e.g., Frenck-Mestre, 2002; Hopp, 2010), it seems more reasonable to understand the differences between L1 and L2 grammatical processing on a continuum that is quantitative in nature than to posit a categorical distinction between L1 and L2 processing mechanisms. Below, we discuss what quantitative differences lie between L1 and L2 sentence processing then.

One quantitative difference that was found in this dissertation is that L2 learners do not seem to apply grammatical constraints as rapidly and effectively as native speakers do. In Experiment 1, the native speakers immediately and completely suppressed the ungrammatical parse that could have been constructed if they heuristically relied on the active gap search strategy, whereas the L2 learners did show evidence that they entertained the illegal gap analysis temporarily, although it was ruled out by the time they made the final plausibility judgment. A possible explanation for L2 learners' delayed application of island constraints is the reduced efficiency in accessing the relevant grammatical knowledge and putting it into use, or in projecting complex syntactic structures that constitute islands. Alternatively, L2 learners may not be able to suppress automatic active gap search as fast and deterministically as native speakers, probably because active gap search is a default, memory-friendly parsing strategy. It could also be explained that the L2 grammatical knowledge is represented in the form that cannot apply to real-time sentence processing as the Shallow Structure Hypothesis postulates. However, as discussed above, the results of this dissertation fit better with the view that the underlying grammatical representations and processes are of a similar nature for L1 and L2 sentence processing.

There was evidence that the generally slower or less efficient processing by L2 learners may not be limited to the application of grammatical constraints but can also characterize filler-gap dependency formation and general structure building. Recall that in Experiment 1, L2 learners seemed to initially ignore island constraints (as shown by a plausibility effect both in the Non-island and Island condition at the region including the first subcategorizer). This appears to suggest that L2 learners do not experience particular difficulty in forming a wh-dependency. In Experiment 3, however, L2 learners seemed to delay filler-gap processing compared with native speakers. Specifically, while native speakers showed sensitivity to the different types of islands immediately at the first subcategorizer in the first-pass regression and regression-path duration, L2 learners showed the corresponding effect two words downstream in the first pass regression. Crucially, L2 learners did not show a main effect of Plausibility at the first subcategorizer unlike in Experiment 1, suggesting that they had not yet considered the filler-gap dependency at the region. This delayed filler-gap dependency evaluation may be due to difficulty in all or some of a series of processes involved in gap processing: identification of a syntactic gap, integration of the filler and the gap, and evaluation of the semantic plausibility of the formulated filler-gap dependency. Putting aside the exact locus of the delay in filler-gap processing, the result suggests that filler-gap dependency formation can also cause difficulty and thus be delayed for L2 learners.

There was also evidence that structure building in general can cause a delay in L2 processing. In Experiment 2, the native speakers experienced more difficulty processing the first subcategorizer in the Island sentences (e.g., *The reporter wondered which book the authors who <u>wrote</u>*.) than in the Non-island sentences (e.g., *The reporter wondered which book the authors* *wrote*.) as shown by the main effect of Island the first-pass regression. This could reflect the processing difficulty when they construct the relative clause modifying the complex NP. L2 learners, however, showed a reversed pattern at the same region in the first pass reading time, taking longer to read the Non-island sentences than the Island sentences. As reflected in the total reading time at the same region, they eventually spent more time reading the Island sentences than the Non-island sentences. The overall pattern of results seems to suggest that it causes greater difficulty for L2 learners than native speakers to construct the relative clause and incorporate it to the preceding NP. Furthermore, the significant interaction of Island and Working Memory in the regression path duration indicates that the degree of difficulty involved in constructing the relative clause was modulated by learners' working memory capacity. The interaction occurred because the higher span learners showed the native-like pattern of taking longer to read the Island sentences than the Non-island sentences, while the lower span learners did not. Available processing resources in L2, therefore, appear to be a factor accounting for the degree of native-likeness in L2 processing.

Another notable quantitative difference between the native speakers and L2 learners was in the ability to reanalyze the initially-formed, ultimately wrong wh-dependencies when encountering the actual subcategorizer. In Experiments 1 and 2, the Non-island sentences were constructed so that the initially formed wh-dependency with the first subcategorizer turned out to be incorrect and needed to be reanalyzed at the actual subcategorizer. Facing this task, in both Experiments 1 and 2, the native speakers showed evidence suggesting that they attempted reanalysis (there was a reversed plausibility effect at the region following the actual subcategorizer), but the L2 learners showed no evidence of attempting reanalysis. Therefore, L2 learners seem to have a particularly severe difficulty in reanalyzing an initial parse. This finding is consistent with the previous finding that L2 learners experience greater difficulty and often fail to reanalyze the initially formed wh-dependencies (Dussias & Piñar, 2010; Williams et al., 2001). However, there are also previous studies suggesting that L2 learners indeed succeed under certain circumstances and quantitative factors influence this success. Roberts and Felser (2011), for example, found that Greek learners of English were generally successful in recovering from the initial misanalysis when processing sentences causing 'weak' garden paths (e.g., *The inspector warned the boss would destroy very many lives*.), while they were less successful in the reanalysis of sentences with 'stronger' garden paths (e.g., *While the band played the song pleased all the customers*.) (cf., Sturt, Pickering, & Crocker, 1999). In Dussias and Piñar (2010), L2 learners with higher working memory capacity conducted reanalysis in a way similar to native speakers in filler-gap processing. This gradient effect of reanalysis complexity, however, does not seem to be limited to L2 learners, because it has also been shown that native speakers of English experience increased difficulty with the more complex reanalysis (e.g., Sturt et al., 1999) and their working memory capacity correlates with the difficulty they experience when processing garden-path sentences (e.g., O'Rourke, 2013). Therefore, the reduced ability to reanalyze the initially-formed wh-dependencies for the L2 learners would be again better explained in terms of quantitative differences between L1 and L2 processing rather than in terms of some sort of categorical distinction associated with different kinds of reanalysis.

As discussed so far, the differences between L1 and L2 speakers can be generally characterized by slower or less efficient processing by L2 learners in applying grammatical constraints, constructing detailed syntactic structures, forming filler-gap dependencies and in reanalyzing the initially-formed filler-gap dependencies. There was also evidence that the speed and efficiency of processing are constrained by the availability of processing resources indicated by reading span. Although late L2 learners are generally slower than native speakers in structural (and semantic) computation as well as in application of grammatical constraints, they still seem to be sensitive to the same set of linguistic factors in ways that are *qualitatively* comparable to native speakers. More specifically, the overall results suggest that late L2 learners can in principle construct as detailed syntactic structures as required for the application of complex grammatical constraints such as island constraints and constraints on parasitic gap distribution during online L2 sentence processing. Thus the findings of this dissertation seem to suggest that the differences between L1 and L2 sentence processing are more quantitative than qualitative.

The present findings do not seems to fit well with the Shallow Structure Hypothesis, according to which the differences between L1 and L2 sentence processing cannot be attributed to the quantitative factors such as L2 learners' slower speed of processing and limited processing resources in the target language. The present findings are rather in line with the views that L1-L2 differences in processing are better explained along quantitative variables such as processing efficiency (e.g., Hopp, 2010) and proficiency in the target language (e.g., Hopp, 2006), and the amount and type of exposure to the target language (e.g., Dussias, 2003; Frenck-Mestre, 2002, 2005). The reliable effect of working memory capacity on the learners' sensitivity to syntactic complexity and the numerical trend showing that the higher span learners applied island constraints faster than the lower span learners suggest that L2 learners' working memory capacity in the L2 may need to be added to the array of variables along which we can compare L1 and L2 processing.

As mentioned earlier with regard to the effect of working memory capacity in L2 learners' sensitivity to structural computation, there was evidence in Experiment 2 that L2 learners' working memory capacity modulates their online sensitivity to the complexity of the syntactic structure to be constructed. The result provides implications for the role of working memory capacity in L2 sentence processing in general. Recall that at the first subcategorizer in Experiment 2, the learners with larger working memory capacity produced a longer regression path duration in the Island conditions than in the Non-island conditions, while those with smaller working memory capacity did not. We interpreted this result as suggesting that only higher working memory learners were successful in constructing the relative clause structure at the processing stage measured by the regression path duration. This result is consistent with the previous finding that L2 learners' success in syntactic analysis of complex structures can be constrained by their working memory capacity (Havik et al, 2009). In Havik et al.'s (2009) self-paced reading study, German L2 learners of Dutch with larger working memory capacity showed evidence for the greater processing difficulty for the object than subject relative clauses that has been well established in the L1 processing literature (e.g., Frazier & Flores d'Arcais, 1989; Kaan, 1997), while the learners with smaller working memory capacity failed to show such an effect. The finding, together the finding of this dissertation, undermine the results of the previous studies that showed a null working memory capacity effect on syntactic computation in L2 (e.g., Roberts & Felser, 2007), based on which the Shallow Structure Hypothesis proposes that L1-L2 differences in sentence processing cannot be attributed to differences in processing resources available to L1 and L2 speakers. In addition, the findings complement the previous studies that showed that working memory capacity modulates L2 learners' processing of plausibility information (e.g., Dussias & Piñar, 2010) and high-level comprehension of L2 text (e.g., Alptekin & Erçtin, 2010; Walter, 2004), suggesting that working memory capacity is one of the factors that affect a wide range of L2 performances.

The discussions made so far regarding the application of island constraints in L2 processing are based on the assumption that the island phenomenon in filler-gap dependency formation is derived from grammatical, not from processing constraints. If the island constraints are not part of the grammar but due to purely cognitive constraints arising from processing complexity of island constructions, the results of this dissertation would be more relevant to the effects of increased processing complexity on wh-dependency formation in L2 than to L2 learners' structural or grammatical processing. The overall results of this dissertation, however, seem to be more consistent with the view that island constraints are represented as part of the grammar, offering some implications for the ongoing debate over the source of the island phenomenon (Phillips, 2012). Most relevant are the results of Experiment 2 that native speakers with larger working memory capacity tended to be *more* sensitive to the island constraints than those with smaller working memory capacity. This pattern of results conflicts the logical prediction of the processing-based accounts of the island phenomenon that readers with larger working memory capacity would be more likely to manage to search for a gap within an island construction (that is, they should be *less* sensitive to the island constraint). The L2 learners in Experiment 2 showed a similar pattern, although the effect was not statistically robust. These results can be easily explained if we assume that the island phenomenon arises due to grammatical constraints and a larger working memory capacity allows greater sensitivity to various types of information (e.g., Just & Carpenter, 1992; Pearlmutter & MacDonald, 1995), including grammatical constraints.

Another piece of evidence for the grammatical status of the island constraints comes from the finding that the native speakers were generally more sensitive to the constraints or faster in applying the constraints than the L2 learners. If the island phenomenon is an epiphenomenon emerging from the processing difficulty involved with complex syntactic structures, it should be more likely for native speakers to posit a gap within an island than for L2 learners because the former will be less vulnerable to processing complexity than the latter, as pointed by Cunnings et al. (2010). Contrary to this prediction of the processing-based account, the native speakers showed the island effect earlier than the L2 learners in Experiment 1 and the native speakers with larger working memory capacity showed sensitivity to island constraints earlier than the L2 learners as a group in Experiment 2. These results will also be explained more easily if the island phenomenon is the result of grammatical constraints, because L2 learners would be in the process of learning grammatical constraints and thus be slower in applying those constraints than native speakers. The present findings thus lend support to the view that the island constraints on long-distance wh-dependencies should be understood as grammar-based phenomena.

Finally, this dissertation contributes to the issue of UG accessibility in adult L2 acquisition. Island constraints have been one of the most investigated area of grammar in relation to this issue. The island constraints have been considered an appropriate test case because they are "obscure and abstract, and they are a parade case of a linguistic phenomenon that is likely to be difficult to observe in the input" (Phillips, 2013). Moreover, since it has been a standard assumption in the L2 acquisition research that islands do not constrain movement in some wh-in-situ languages, the constraints have been deemed adequate for the purpose of examining how the non-existence of certain grammatical features in the L1 influences acquisition of those features in the L2. Studies that examined whether L2 learners from the wh-in-situ language background can learn the island constraints so far have produced mixed findings. The existing data, however, is leaning in favor of the view that adult learners from the wh-in-situ language L1 are not able to develop native-like knowledge of island constraints (e.g., Bley-Vroman, Felix, & loup, 1988;

Hawkins & Chan, 1997; Hawkins & Hattori, 2006; Johnson & Newport, 1991; Schachter, 1989, 1990) than in favor of the opposite view (e.g., White & Juffs, 1998). The finding of the present study adds to relatively fewer studies that suggest that adult L2 learners from wh-in-situ language backgrounds can indeed learn the island constraints in the L2, by showing that adult learners with a wh-in-situ L1 (i.e., Korean) can develop sensitivity to island constraints as well as subtle grammatical knowledge about distribution of parasitic gaps, even to the extent that the constraints can be applied in real-time processing.

To conclude, the results of this dissertation provide some important implications for multiple issues including the source of the island phenomenon, the nature of L2 syntactic processing and grammatical knowledge, and UG accessibility in late L2 acquisition. Despite the fact that the island constraints operate based on quite complex syntactic structures and are hard to learn from linguistic input (cf, Pearl & Sprouse, 2013), this dissertation offers evidence that advanced L2 learners are sensitive to the constraints in the course of filler-gap processing, even under a natural reading condition where explicit attention to grammatical aspects is not required. At the same time, the results also suggest that L2 sentence processing in general and application of island constraints in particular were not as rapid or automatized as compared to in L1 sentence processing. The overall results seem to undermine the view that if a second language is not fully acquired within a maturationally constrained optimal period, it cannot be learned and processed through the same kind of mechanisms as those underlying L1 acquisition and sentence processing (Bley-Vroman, 1990; Clahsen & Felser, 2006a, b, c).

References

Aldwayan, S., Fiorentino, R., & Gabriele, A. (2010). Evidence of syntactic constraints in the processing of wh-movement: A study of Najdi Arabic learners of English. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 65-86). Amsterdam: John Benjamins.

Alptekin, C., & Erçtin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading, 33*, 206–219.

Aoshima, Sachiko, Colin Phillips & Amy S. Weinberg. 2004. Processing filler–gap dependencies in a head-final language. *Journal of Memory and Language, 51*, 23–54.

Aoun, J., & Li, Y. H. (1993). On some differences between Chinese and Japanese. *Linguistic Inquiry*, *24*, 365-372.

Belikova, A., & White, L. (2009). Evidence for the fundamental difference hypothesis or not?: Island constraints revisited. *Studies in Second Language Acquisition*, *31*, 199-223.

Berwick, R., & Weinberg, A. S. (1984). *The grammatical basis of linguistic performance*. Cambridge, MA: MIT Press.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*, 3-49.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*, 175–198.

Bley-Vroman, R. W., Felix, S. W., & loup, G. L. (1988). The accessibility of Universal Grammar in adult language learning. *Second Language Research, 4*(1), 1-32.

Boland, J., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb-argument structure in parsing and interpretation: Evidence from wh-Questions. *Journal of Memory and Language, 34*, 774–806.

Bourdages, F. S. (1992). Parsing complex NPs in French. In H. Goodluck & M.S. Rochemont (Eds,), *Island constraints: Theory, acquisition and processing* (pp. 61-87). Dordrecht: Kluwer.

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*, 2492-2511.

Carroll, S. E. (1999). Putting ‘input’ in its proper place. *Second Language Research, 15*, 337–388.

Choe, J. (1987). LF movement and pied-piping. *Linguistic Inquiry*, *18*(2), 348-353.

Chomsky, N. (1965). *Aspect of the theory of syntax*. Cambridge : The MIT Press.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 232-286). New York: Holt, Rinehart and Winston.

Chomsky, N. (1986). *Barriers*. Cambridge : The MIT Press.

Chomsky, N. (1977). On wh-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal Syntax* (pp. 71-132). New York: Academic Press, Inc.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Christianson, K., Hollingworth, A. Halliwell, J. and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*. 368–407.

Cinque, G. (1991). *Types of A’-dependencies*. Cambridge, MA: MIT Press.

Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3-42.

Clahsen, H., & Felser, C. (2006b). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*, 107-126.

Clahsen, H., & Felser, C. (2006c). How native-like is non-native language processing? *TRENDS in Cognitive Sciences, 10*(12), 564-570.

Clifton, C. Jr., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In M.K. Tanenhaus & G.N. Carlson (Eds.), *Linguistic structure in language processing* (pp. 273-317). Dordrecht: Kluwer.

Clifton, C. Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 341-372). Oxford: Elsevier Ltd.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Karttunen & M. Zwicky (Eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives* (pp. 94-128). Cambridge: Cambridge University Press.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O. & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769–786.

Culicover, P., & Postal, P. (Eds.) (2001). *Parasitic gaps*. Cambridge, MA: MIT Press.

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2010). Constraints on L2 learners' processing of wh-dependencies: Evidence from eye movements. In B. VanPatten & J. Jegerski (Eds.), *Research in Second Language Processing and Parsing* (pp. 87-110). Amsterdam: John Benjamins.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in comprehending and producing words in context. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466.

Deane, P. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics, 2*, 1-63.

Dekydtspotter, L., Schwartz, B. D., Sprouse, R. A. (2006). The comparative fallacy in L2 processing research. In M. O'Brien, C. Shea & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 33-40). Cascadilla Press, Somerville, MA.

de Vincenzi, M. (1991). *Syntactic parsing strategies in Italian*. Dordrecht:Kluwer.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, *25*, 529–557.

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*, 501–513.

Dussias, P., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of *wh*-gaps by Chinese–English second language speakers. *Second Language Research*, *26*(4), 443-472.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English L2 speakers. *Bilingualism: Language and Cognition*, *10*, 101–116.

Engdahl, E. (1983). Parasitic gaps. *Linguistics and Philosophy*, *5*, 5-34.

Felser, C., Cunnings, I. Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition, 34*(1), 67-98.

Felser, C., & Roberts, L. (2007). Processing wh–dependencies in a second language: A cross–modal priming study. *Second Language Research*, *23*, 9–36.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*, 453–489.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47,* 164–203.

Ferreira, F., & Patson, N. D. (2007). The 'good-enough' approach to language comprehension. *Language and Linguistics Compass*, *1*, 71-83.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *31*, 725-745.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, *9,* 427-473.

Fodor, J. D. (1983). Phrase structure parsing and the island constraints. *Linguistics and Philosophy*, *6*, 163-223.

Freedman, S., & Forster, K. (1985). The psychological status of overgenerated sentences. *Cognition*, *19*, 101–131.

Francis, W., & Kucera, H. (1982). *Frequency Analysis of English Usage*. New York: Houghton Mifflin.

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory, 5*, 519–559.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

Frazier, L., & Flores d'Arcais, G. (1989). Filler driven parsing: A study of gap filling in Dutch. Journal of Memory and Language, 28, 331–344.

Frenck-Mestre, C. (1997). Examining second language reading: An on-line look. In A. Sorace, C. Heycock & R. Shillcock (Eds.), *Language acquisition, knowledge representation and processing* (pp. 474-478). Holland: Emerald Group Publishing.

Frenck-Mestre, C. (2002). An on–line look at sentence processing in the second language. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217–236). New York: Elsevier.

Frenck-Mestre, C. (2005). Eye-movement as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, *21*, 175–198.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Quarterly Journal of Experimental Psychology*, *50A*, 119–148.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 51–60.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1−76.

Gibson, E., & Warren, T. (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*(1), 55–78

Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammar: A prosodic approach. *Second Language Research, 22*, 243-268.

Hammerton, J., Osborne, M., Armstrong, S., & Daelemans, W. (2002). Introduction to special issue on machine learning: Approaches to shallow parsing. *Journal of Machine Learning Research*, *2*, 551-558.

Han, C.-H., & Kim, J.-B. (2004). Are there "double relative clauses" in Korean? *Linguistic Inquiry*, *35*, 315-337.

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*(1), 73-112.

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across languages. *Language*, *75,* 224-285.

Hawkins, R. (2001). *Second Language Syntax: A Generative Introduction*. Oxford: Blackwell Publishers.

Hawkins, R., & Chan, Y.-H. C. (1997). The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. *Second Language Research*, *13*, 187-226.

Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research*, *22*, 269-301.

Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints on syntactic islands. *Language*, *86*, 366-415.

Hoover, M. L., & Dwivedi, V. D. (1998). Syntactic processing by skilled bilinguals. *Language Learning*, *48*(1), 1-29.

Hopp, H. (2006). Syntactic features and reanalysis in near–native processing. *Second Language Research*, *22*, 369–397.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*, 901–931.

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar*. Unpublished Ph.D. dissertation, MIT, Cambridge.

Ionin, T., & Montrul, S. (2010). The role of L1-transfer in the interpretation of articles with definite plurals in L2-English. *Language Learning*, *60*(4), 877-925.

Ionin, T., Baek, S., Kim, E., Ko, H., & Wexler, K. (2012). *That*'s not so different from *the*: Definite and demonstrative descriptions in second language acquisition. *Second Language Research*, *28*(1), 69-101.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*, 603-634.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, *39*(3): 215-258.

Juffs, A. (2001). Psycholinguistically oriented second language research. *Annual Review of Applied Linguistics*, *21*, 207–20.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, *102*(2), 199-225.

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition*, *17*, 483-516.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149.

Kaan, E. (1997). Processing subject-object ambiguities in Dutch. Groningen Dissertations in Linguistics 20.

Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. In P. W. Culicover & L. McNally (Eds.), *Syntax and Semantics 29: The Limits of Syntax* (pp. 241-279). Academic Press.

Kluender, R. (2004). Are subject islands subject to a processing account? In B. Schmeiser, V. Chand, A. Kelleher & A.J. Rodriguez (Eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics Somerville* (pp. 101-125). MA: Cascadilla Press.

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, *8*, 573-633.

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14*, 359–375.

Lasnik, H., & Saito, M. (1984). On the nature of proper government. *Linguistic Inquiry*, *15*(2), 235-289.

Lasnik, H. & Saito, M. (1992). *Move alpha: Conditions on its application and output*. Cambridge, MA: MIT Press.

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, *33*, 51-73.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Manzini, M. R. (1992). *Locality: A theory and some of its empirical consequences*. Cambridge, MA: MIT Press.

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, *19*(2), 144-161.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*, 53–78.

Martohardjono, G., & Gair, J. (1993). Apparent UG inaccessibility in second language acquisition: Misapplied principles or principled misapplications? In F. Eckman (Ed.), *Confluence: Linguistics, L2 acquisition and speech pathology* (pp. 79-103). Amsterdam: John Benjamins.

McDonald, M. C. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*, 381-401.

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 432-460.

McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognitive Processes*, *11*, 495-523.

McLaughlin, J., Tanner, D., Pitkanen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*(Suppl. 2), 123–150.

Mitchell, D. C., & Cuetos, F. (1991). The origins of parsing strategies. Conference proceedings: Current issues in natural language processing (pp. 1-12). Austin: University of Texas at Austin, CT.

Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research*, *21*(2), 152-174.

Neville, H., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically-based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*,151–65.

O'Neill, R., Cornelius, E., & Washburn, G. (1981). *American Kernel Lessons: Advanced Students' Book*. New York: Longman Publishing Group.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*, 563-588.

O'Rourke, P. (2013). The interaction of different working memory mechanisms and sentence processing: A study of the P600. In M. Knauff, N. Sebanz, M. Pauen, I. Wachsmuth. (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1097-1102). Austin, TX: Cognitive Science Society.

Pan, H.-U., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*, 221-236.

Papadopoulou, D. (2005). Reading-time studies of second language ambiguity resolution. *Second Language Research*, *21*, 98–120.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: a study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *24*, 501–528.

Pearlmutter, N. J. , Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*, 427-456.

Pearlmutter, N. J., & MacDonald, M. E. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, *34*, 521–542.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*, 795–823.

Phillips, C. (2013). On the nature of island constraints. I: Language processing and reductionist accounts. In J. Sprouse & N. Hornstein (Eds.), *Experimental syntax and island effects* (pp. 64-108). Cambridge University Press.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the Interfaces, Syntax and Semantics*, (vol.37), (pp. 153-186). Bingley, UK: Emerald.

Pickering, M. J., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 199-224). London: Lawrence Erlbaum.

Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes, 18*, 469-503.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language? Evidence from Tense and Agreement. Second Language Research, 16, 103-133.

Pritchett, B. L. (1991), Subjacency in a principle-based parser. In R.C. Berwick, S.P. Abney & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 301-345). Dordrecht: Kluwer.

Pritchett, B. L. (1992). Parsing with grammar: Islands, heads, and garden paths. In H. Goodluck & M. Rochemont (Eds.), *Island constraints: Theory, acquisition, and processing* (pp. 321-349). Dordrecht: Kluwer Academic.

Rah, A., & Adone, D. (2010). Relative clause versus main verb ambiguity in L2 learners at different proficiency levels. *Studies in Second Language Acquisition*, *32*, 79–109.

Rizzi, L. (1990). *Relativized Minimality*. Mass: MIT Press.

Rizzi, L. (2001). Relativized minimality effects. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory, 4*, 89-110. Oxford: Blackwell.

Roberts, L. (2012). Psycholinguistic techniques and resources in SLA research. *Second Language Research*, *28*(1), 113-127.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*, 299–331.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at gap positions in children's sentence processing. *Journal of Psycholinguistic Research*, *36*, 175–188.

Ross, J. R. (1967). *Constraints on variables in syntax.* Ph.D. dissertation, MIT.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *TRENDS in Cognitive Sciences*, *6*(9), 382-386.

Schachter J (1989) Testing a proposed universal. In Gass S and Schachter J (eds) *Linguistic perspectives on second language acquisition*. Cambridge: Cambridge University Press, 73-88.

Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research, 6*(2), 93-124.

Schachter, J. & Yip, V. (1990). Grammaticality judgments: Why does anyone object to subject extraction? *Studies in Second Language Acquisition*, *12*, 379–392.

Schwartz , B. D., & Sprouse , R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*, 40–72 .

Stowe, L. A. (1986). Evidence for on-line gap-location. *Language and Cognitive Processes, 1*, 227-245.

Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language & Cognitive Processes*, 18, 143-161.

Szabolcsi, A., & den Dikken, M. (1999). Islands. *GLOT International, 4*, 3-9.

Szabolcsi, A., & Zwarts, F. (1993). Weak islands and an algebraic semantics for scope taking. *Natural Language Semantics, 1*, 235-285.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis diculty in language comprehension. *Journal of Memory and Language, 40*(1), 136-150.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50*, 355-370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition,16*, 367-382.

Traxler, M. J., & Pickering, M. J. (1996) Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35: 454-475.

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition,* 33, 339-372.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(1), 105-122.

Walter, C. (2004). Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. Applied Linguistics, 25, 315–339.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. Journal of *Cognitive Neuroscience, 8*, 231-256.

Weber-Fox, C. M., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed-class words: an ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research, 44*, 1338-1353.

White, L. (1992). Subjacency violations and empty categories in L2 acquisition. In H. Goodluck & M. Rochemont (Eds.), *Island Constraints* (pp. 445-464). Dordrecht: Kluwer.

White, L. (2003). *Second Language Acquisition and Universal Grammar.* Cambridge: Cambridge University Press.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12*(3), 233-265.

White, L., & Juffs, A. (1998). Constraints on wh-movement in two different contexts of nonnative language acquisition: Competence and processing. In S. Flynn, G. Martohardjono & W.A. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 111-129). Mahwah, NJ: Lawrence Erlbaum.

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition, 9*(1), 71–88.

Williams, J. N., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics, 22*, 509-540.

Yoshida, M. (2006). *Constraints and mechanisms in long-distance dependency formation*. PhD dissertation, University of Maryland.

Yoon, J.-M. (2006). Selective island effects in Korean wh-constructions: A diachronic approach. *Korean Journal of Linguistics*, *31*(2), 301-340.

Appendix A
Materials for Experiments 1, 2 and 3

Experiment 1
Materials for Stop-making-sense task

The sentences in the first item are presented as they were in the experiment. In the following items, of the alternative NPs in the relative clause subject position, the left of the slash bar represents the plausible condition and the right represents the implausible condition (e.g., *hotel/family*). In the actual subcategorizer position, the left represents the non-island condition and the right, paired with the relative pronoun *who* in parentheses, represents the island condition (e.g., *for*/(*hated*)). When the prepositions in the non-island condition differ between the plausible and implausible condition, the two prepositions were juxtaposed with a slash in between (e.g., *for/behind/*(*saw*)).

Non-island/Plausible condition
I remember which hotel the maid cleaned carelessly for before she was fired.
Non-island/Implausible condition
I remember which family the maid cleaned carelessly for before she was fired.
Island/Plausible condition
I remember which hotel the maid who cleaned carelessly hated before she was fired.
Island/Implausible condition
I remember which family the maid who cleaned carelessly hated before she was fired.

I remember which soundtrack/Olympics the singer (who) sang beautifully for/(advertised) when she was at the peak of her career.

I know which hospital/detergent the janitor (who) cleaned thoroughly at/(liked) because I've known him for a long time.

I wonder which animal/knife the king (who) hunted frequently for/(liked) when he went on hunting trips.

I wonder which flute/racket the child (who) played cheerfully with/(wanted) when he was in kindergarten.

I forgot which war/year the army (who) lost badly in/(recalled) before I watched the war movie again.

I asked which magazine/novelist the housewife (who) read occasionally about/(knew) when I interviewed her for local news.

I learned which friend/lamp the man (who) killed intentionally with/(used) when he wanted to get revenge.

I asked which theory/museum the student (who) read frequently about/(liked) while she was studying in college.

I know which story/river the reporter (who) wrote enthusiastically about/(liked) before he died of cancer last month.

I wonder which cloth/soap the traveler (who) washed infrequently with/(used) after he lost his bag.

I learned which beans/oil the woman (who) cooked skillfully with/(bought) when I took a cooking lesson from her.

I know which harmonica/blocks the boy (who) played happily with/(broke) while he was in his room.

I asked which deer/tree the man (who) hunted stealthily for/behind/(saw) when we visited his favorite hunting place.

I asked which lottery/table the gambler (who) won recently in/at/(remembered) when I interviewed him.

I wonder which meat/spoon the chef (who) cooked creatively with/(used) when he prepared the exotic food.

I forgot which opera/theater the woman (who) sang elegantly in/(preferred) when she worked as a professional soprano.

I remember which general/country the spy (who) killed mercilessly for/after/(betrayed) after he was paid a lot of money.

I wonder which towel/cleanser the girl (who) washed carefully with/(used) while she was camping in the mountains.

I asked which book/city the writer (who) wrote passionately about/(saw) while he was traveling.

Experiment 1
Materials for the offline grammaticality judgment task

The first of a pair of sentences are grammatical, the second ungrammatical.

Grammatical condition
I asked which publisher the author who wrote the book passionately saw while he was traveling.
Ungrammatical condition
I asked which book the author who wrote passionately saw the publisher while he was traveling.

I know which editor the reporter who wrote the story enthusiastically liked before he died of cancer last month.
I know which story the reporter who wrote enthusiastically liked the editor before he died of cancer last month.

I asked which library the student who read the theory frequently liked while she was studying in college.
I asked which theory the student who read frequently liked the library while she was studying in college.

I asked which journalist the housewife who read the magazine occasionally knew when I interviewed her for local news.
I asked which magazine the housewife who read occasionally knew the journalist when I interviewed her for local news.

I forgot which theater the woman who sang the opera elegantly preferred when she worked as a professional soprano.
I forgot which opera the woman who sang elegantly preferred the theater when she worked as a professional soprano.

I remember which movie the singer who sang the soundtrack beautifully advertised when she was at the peak of her career.
I remember which soundtrack the singer who sang beautifully advertised the movie when she was at the peak of her career.

I wonder which spoon the chef who cooked the meat creatively used when he prepared the exotic food.
I wonder which meat the chef who cooked creatively used the spoon when he prepared the exotic food.

I learned which oil the woman who cooked the beans skillfully bought when I took a cooking lesson from her.
I learned which beans the woman who cooked skillfully bought the oil when I took a cooking lesson from her.

I wonder which detergent the girl who washed the towel carefully used while she was camping in the mountains.
I wonder which towel the girl who washed carefully used the detergent while she was camping in the mountains.

I wonder which soap the traveler who washed the cloth infrequently used after he lost his bag.
I wonder which cloth the traveler who washed infrequently used the soap after he lost his bag.

I know which doctor the janitor who cleaned the hospital thoroughly liked because I've known him for a long time.

I know which hospital the janitor who cleaned thoroughly liked the doctor because I've known him for a long time.

I remember which manager the maid who cleaned the hotel carelessly hated before she was fired.
I remember which hotel the maid who cleaned carelessly hated the manager before she was fired.

I wonder which arrow the king who hunted the animal frequently liked when he went on hunting trips.
I wonder which animal the king who hunted frequently liked the arrow when he went on hunting trips.

I asked which tree the man who hunted the deer stealthily saw when we visited his favorite hunting place.
I asked which deer the man who hunted stealthily saw the tree when we visited his favorite hunting place.

I know which tooth the boy who played the harmonica happily broke while he was in his room.
I know which harmonica the boy who played happily broke the tooth while he was in his room.

I wonder which toy the child who played the flute cheerfully wanted when he was in kindergarten.
I wonder which flute the child who played cheerfully wanted the toy when he was in kindergarten.

I remembered which country the spy who killed the general mercilessly betrayed after he was paid a lot of money.
I remembered which general the spy who killed mercilessly betrayed the country after he was paid a lot of money.

I learned which plot the man who killed the friend intentionally used when he wanted to get revenge.
I learned which friend the man who killed intentionally used the plot when he wanted to get revenge.

I asked which number the gambler who won the lottery recently remembered when I interviewed him.
I asked which lottery the gambler who won recently remembered the number when I interviewed him.

I forgot which year the army who lost the war badly recalled before I watched the war movie again.
I forgot which war the army who lost badly recalled the year before I watched the war movie again.

Experiment 2

Non-island/Plausible
The police learned which hotel the maid cleaned occasionally for before she disappeared.
Non-island/Implausible
The police learned which family the maid cleaned occasionally for before she disappeared.
Island/Plausible
The police learned which hotel the maid who cleaned occasionally sued before she disappeared.
Island/Implausible
The police learned which family the maid who cleaned occasionally sued before she disappeared.

The wife heard which building/detergent the janitor (who) cleaned thoroughly at/with/(liked) because it was new.

The vegetarian watched which beans/cutter the chef (who) cooked skillfully with/(cooked) last evening on TV.

The judges watched which meat/tools the chefs (who) cooked creatively with/(used) during the cooking contest.

The boys heard which animal/location their father (who) hunted stealthily for/at/(killed/used) a long time ago.

The duke knew which animals/knife the king (who) hunted frequently for/with/(liked) when he went on hunting trips.

The CIA learned which general/country the spy (who) killed mercilessly for/(betrayed) after being paid much money.

The news reported which gang/knife the man (who) killed brutally for/with/(hired/used) early this morning .

The movie described which election/year the politician (who) won greatly in/(recalled) alongside his supporters.

The dad heard which musical instrument/neighborhood friend the boy (who) played happily with/(liked) yesterday over the phone.

The parents heard which toy piano/family pet the child (who) played cheerfully with/(liked) today in his room.

The mother heard which letters/teacher her young son (who) read curiously about/(liked) yesterday in kindergarten.

The teacher recalled which magazine/library the student (who) read frequently about/(edited) while in high school.

The tenor recalled which operas/theaters the soprano (who) sang elegantly in/(reviewed) before she died.

The queen recalled which story/garden the poet (who) sang beautifully about/in/(loved) before he died of illness.

The friends heard which cloth/fountain the traveler (who) washed frequently with/in/(used) while visiting the desert city.

The girl watched which towel/soap her mother (who) washed carefully with/(used) after applying cleansing cream.

The newspaper reported which lottery/table the gambler (who) won recently in/at/(recommended) this morning.

The editor heard which story/laptop the reporter (who) wrote enthusiastically about/(liked) yesterday at a meeting.

The reporter asked which book/city the author (who) wrote passionately about/(saw) while he was traveling.

Experiment 3

Subject with an infinitival complement/Plausible
Which items did the plan to prepare carefully for the party include on the list?
Subject with an infinitival complement/Implausible
Which mistakes did the plan to prepare carefully for the party avoid on her birthday?
Subject with a relative clause modifier/Plausible
Which items did the plan that prepared carefully for the party include on the list?
Subject with a relative clause modifier/Implausible
Which mistakes did the plan that prepared carefully for the party avoid on her birthday?

Which students/errors did the plan to/(that) prepare(d) thoroughly for the exam motivate/prevent in the class?

Which supplies/accident did the policy to/(that) prepare(d) thoroughly for the snow storm require/prevent in the winter?

Which people/election did the campaign to/(that) warn(ed) firmly against drugs influence throughout the nation?

Which type of drivers/accidents did the campaign to/(that) warn(ed) seriously against drunk driving affect/reduce last year?

Which students/policy did the campaign to/(that) warn(ed) strongly against cigarettes scare/create at school?

Which citizens/obstacles did the plan to/(that) help(ed) financially with medical bills save/solve in the nation?

Which manager/complaints did the idea to/(that) help(ed) effectively with table setting please/reduce at the restaurant?

Which person/food did the plan to/(that) help(ed) effectively with weight loss please/eliminate at the gym?

Which classes/complaints did the plan to/(that) improve(d) significantly on the lecture enhance/avoid during the semester?

Which treatment/viruses did the effort to/(that) improve(d) thoroughly on the vaccine enable/block in the hospital?

Which city/tax did the plan to/(that) improve(d) considerably on highway conditions benefit/increase last year?

Which client/law did the proposal to/(that) compensate(d) fully for the lost money satisfy/follow during the meeting?

Which customer/lawsuit did the plan to/(that) compensate(d) quickly for the mistake pacify/prevent in the company?

Which relatives/arguments did the plan to/(that) compensate(d) partially for the divorce anger/cause in the family?

Which religious group/improvements could the campaign to/(that) fight/(fought) persistently for civil rights benefit/create in society?

Which dictator/benefits did the plan to/(that) fight/(fought) bitterly for democracy overthrow/produce in the past decade?

Which disease/drinks did the plan to(that) fight/(fought) effectively against infection prevent/eliminate in the diabetic patient?

Which schools/students will the proposal to/(that) expand(ed) greatly on the curriculum advance/overwhelm next year?

Which park/painters did the idea to/(that) expand(ed) greatly on the architect's design modernize/employ in the city?

Appendix B
Cloze test

For each blank in the following passage, please circle one of three options given. Please choose the option appropriate for the context. Please choose one option only for each blank.

Joe came home from work on Friday. It was payday, but he wasn't **(1) even / more / ever** excited about it. He knew that **(2) then / when / while** he sat down and paid his **(3) checks / bills / salary** and set aside money for groceries, **(4) driving / pay / gas** for the car and a small **(5) deposit / withdrawal / money** in his savings account, there wouldn't be **(6) quite / not / too** much left over for a good **(7) pleasure / leisure / life**.

He thought about going out for **(8) eat / dinner / eating** at his favorite restaurant, but he **(9) just / only / very** wasn't in the mood. He wandered **(10) around / at / in** his apartment and ate a sandwich. **(11) In / For / After** a while, he couldn't stop himself **(12) for / from / about** worrying about the money situation. Finally, **(13) he / she / it** got into his car and started **(14) drive / driven / driving**.

He didn't have a destination in **(15) head / mind / fact**, but he knew that he wanted **(16) be / to be / being** far away from the city **(17) which / there / where** he lived. He turned onto a quiet country **(18) road / house / air**. The country sights made him feel **(19) as good / better / best**. His mind wandered as he drove **(20) past / in / to** small farms and he began to **(21) try / think / imagine** living on his own piece of **(22) house / land / farm** and becoming self-sufficient. It had always **(23) being / been / be** a dream of his, but he **(24) having / have / had** never done anything to make it **(25) a / one / some** reality. Even as he was thinking, **(26) their / his / her** logical side was scoffing at his **(27) favorite / practical / impractical** imaginings. He debated the advantages and **(28) cons / disadvantages / problems** of living in the country and **(29) growing / breeding / building** his own food. He imagined his **(30) farmhouse / truck / tractor** equipped with a solar energy panel **(31) at / out / on** the roof to heat the house **(32) in / for / over** winter and power a water heater. **(33) She / He / They** envisioned fields of vegetables for canning **(34) either / and / but** preserving to last through the winter. **(35) Whether / Even / If** the crops had a good yield, **(36) maybe / possible / may** he could sell the surplus and **(37) store / save / buy** some farming equipment with the extra **(38) economy / cost / money**.

Suddenly, Joe stopped thinking and laughed **(39) at / out / so** loud, "I'm really going to go **(40) through / away / in** with this?"

Adapted from American Kernel Lessons: Advanced Students' Book. O'Neill, Cornelius and Washburn (1981).